IDA PAPER P-2241

THE ECONOMIC BENEFITS OF PREDICTING JOB PERFORMANCE

Joseph Zeidner
Cecil D. Johnson

with chapters contributed by
Edward Schmitz
Roy Nord
U.S. Army Research Institute

September 1989

Prepared for
Office of the Under Secretary of Defense for Acquisition
(Research and Advanced Technology)

IDA
INSTITUTE FOR DEFENSE ANALYSES
1801 N. Beauregard Street, Alexandria, Virginia 22311-1772

DEFINITIONS

IDA publishes the following documents to report the results of its work.

Reports

Reports are the most authoritative and most carefully considered products IDA publishes. They normally embody results of major projects which (a) have a direct bearing on decisions affecting major programs, or (b) address issues of significant concern to the Executive Branch, the Congress and/or the public, or (c) address issues that have significant economic implications. IDA Reports are reviewed by outside panels of experts to ensure their high quality and relevance to the problems studied, and they are released by the President of IDA.

Group Reports

Group Reports record the findings and results of IDA established working groups and panels composed of senior individuals addressing major issues which otherwise would be the subject of an IDA Report. IDA Group Reports are reviewed by the senior individuals responsible for the project and others as selected by IDA to ensure their high quality and relevance to the problems studied, and are released by the President of IDA.

Papers

Papers, also authoritative and carefully considered products of IDA, address studies that are narrower in scope than those covered in Reports. IDA Papers are reviewed to ensure that they meet the high standards expected of refereed papers in professional journals or formal Agency reports.

Documents

IDA Documents are used for the convenience of the sponsors or the analysts (a) to record substantive work done in quick reaction studies, (b) to record the proceedings of conferences and meetings, (c) to make available preliminary and tentative results of analyses, (d) to record data developed in the course of an investigation, or (e) to forward information that is essentially unanalyzed and unevaluated. The review of IDA Documents is suited to their content and intended use.

The work reported in this document was conducted under contract MDA 903 84 C 0031 for the Department of Defense. The publication of this IDA Paper does not indicate endorsement by the Department of Defense, nor should the contents be construed as reflecting the official position of that Agency.

This Paper has been reviewed by IDA to assure that it meets the high standards of thoroughness, objectivity, and appropriate analytical methodology and that the results, conclusions and recommendations are properly supported by the material presented.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1989	Final--June 1988 to July 1989

4. TITLE AND SUBTITLE
The Economic Benefits of Predicting Job Performance

5. FUNDING NUMBERS
C - MDA 903 84 C 0031
T - T-D2-435

6. AUTHOR(S)
Joseph Zeidner, Cecil D. Johnson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311

8. PERFORMING ORGANIZATION REPORT NUMBER
IDA Paper P-2241

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
OUSD(A)/R&AT
The Pentagon, Room 3D129
Washington, DC 20301

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report examines the adequacy of the present ASVAB aptitude area composites. A utility analysis provides productivity gains in dollar-valued terms attributable to changes in the ASVAB job entry standards and assignment procedures. Realistic estimates of costs and benefits of alternative manpower selection and classification policies are needed to provide military policymakers with rational choices in allocating scarce resources among strategies.

Using least squares estimates of performance in each job family in place of operational aptitude composites for initial assignment increases mean predicted performance 0.143 standard deviation units over the current selection and assignment process, a present net value gain of over $260 million each year.

Simulation results show that the present aptitude area composites are of limited value, but there is considerable classification efficiency potentially obtainable from the present ASVAB if it is used in accordance with differential assignment principles.

A set of recommendations for proposed changes in the operational use of the ASVAB over a five-year period is made on the basis of simulation results, prior research findings and psychometric theory. Although the analysis was conducted in the Army context, the recommendations are applicable to all services.

A series of ongoing research efforts expressly designed to increase further the potential selection and classification efficiency of the ASVAB are detailed.

14. SUBJECT TERMS
military personnel selection and classification, aptitude testing, job performance, economic benefits, manpower, classification, selection utility, utility analysis

15. NUMBER OF PAGES
272

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

NSN 7540 01-280-5500

Standard Form 298 (Rev 2 89)
Prescribed by ANSI Std Z39 18
298 102

IDA PAPER P-2241

THE ECONOMIC BENEFITS OF PREDICTING JOB PERFORMANCE

Joseph Zeidner
Cecil D. Johnson

with chapters contributed by
Edward Schmitz
Roy Nord
U.S. Army Research Institute

September 1989

INSTITUTE FOR DEFENSE ANALYSES

Contract MDA 903 84 C 0031
Task T-D2-435

ACKNOWLEDGMENTS

We wish to express our sincere appreciation for support of this study by making the Army 1984 recruit accession base available and by arranging for the simulations required for the utility analysis to:

Edgar Johnson, Technical Director, The Army Research Institute (ARI),

N. Kent Eaton, Director, Manpower and Personnel Research Laboratory, ARI,

Curtis Gilroy, Chief, Manpower and Personnel Policy Research Group, ARI.

The efforts of Arthur Drucker, formerly at the Army Research Institute, in editing the manuscript are gratefully acknowledged.

We are grateful for the reviews and useful technical comments made by:

Hubert Brogden, Professor Emeritus, Purdue University

Wayne Cascio, University of Colorado at Denver

Curtis Gilroy, The Army Research Institute

Paul Horst, Professor Emeritus, University of Washington

Edgar Johnson, The Army Research Institute

Robert Sadacca, HumRRO

William Sands, Navy Personnel Research and Development Center

Frank Schmidt, The University of Iowa

Richard Sorenson, Navy Personnel Research and Development Center.

This study was performed for the Office of the Under Secretary of Defense (Acquisition)/(Research and Advanced Technology). Technical cognizance for the work, under Task Order T-D2-435, was assigned to the Assistant for Training and Personnel Technology. We are indebted to Earl Alluisi for his contributions to and interest in the effort, and to Jesse Orlansky, Institute for Defense Analyses, for suggesting that the utility of selection and classification procedures be evaluated, for providing valuable information, and for his many useful suggestions on preparing this report.

ABSTRACT

This report examines the adequacy of the present ASVAB aptitude area composites. A utility analysis provides productivity gains in dollar-valued terms attributable to changes in the ASVAB job entry standards and assignment procedures. Realistic estimates of costs and benefits of alternative manpower selection and classification policies are needed to provide military policymakers with rational choices in allocating scarce resources among strategies.

Using least squares estimates of performance in each job family in place of operational aptitude composites for initial assignment increases mean predicted performance 0.143 standard deviation units over the current selection and assignment process, a present net value gain of over $260 million each year.

Simulation results show that the present aptitude area composites are of limited value, but there is considerable classification efficiency potentially obtainable from the present ASVAB if it is used in accordance with differential assignment principles.

A set of recommendations for proposed changes in the operational use of the ASVAB over a five-year period is made on the basis of simulation results, prior research findings and psychometric theory. Although the analysis was conducted in the Army context, the recommendations are applicable to all services.

A series of ongoing research efforts expressly designed to increase further the potential selection and classification efficiency of the ASVAB are detailed.

CONTENTS

TABLES

FIGURES

ABBREVIATIONS

AA	Aptitude Area
ACB	Army Classification Battery
AFQT	Armed Forces Qualification Test
AGCT	Army General Classification Test
AI	Aptitude Index
AR	Arithmetic Reasoning
AS	Auto Shop Information
ASVAB	Armed Services Vocational Aptitude Battery
AVF	All-Volunteer Force
CL	Clerical/Administrative
CO	Combat
CREPID	Cascio-Ramos Estimate of Performance In Dollars
CS	Coding Speed
DEP	Delayed Entry Program
DI	Disposition Index
EI	Electronic Information
EL	Electronic Repair
EPAS	Enlisted Personnel Allocation System
F	Finger Dexterity
FA	Field Artillery
FLS	Full Least Squares
g	General Component
G	General Intelligence
GATB	General Aptitude Test Battery
GM	General Maintenance

GS	General Science
GVN	Cognitive Ability
H_a	Absolute Validity Index
H_d	Differential Validity Index
HRA	Human Resource Accounting
LSE	Least Squares Estimate
LP	Linear Program
MAXAACON	Maximize Aptitude Area Score, Constrained
MAXAAFREE	Mazimize Aptitude Area Score, Unconstrained (or Full)
Max-PSE	Maximum Potential Selection Efficiency
MC	Mechanical Comprehension
MDS	Multidimensional Screening
MK	Mathematical Knowledge
MM	Mechanical Maintenance
MOS	Military Occupational Speciality
MOSLS	Military Occupational Speciality Level System
MPP	Mean Predicted Performance
N	Numerical Aptitude
NO	Numerical Operations
OAE	Operational Allocation Efficiency
OF	Operators/Food
OPM	Office of Personnel Management
OPTAACL	Optimization on Aptitude Area Score, Classification Only
OPTAASC	Optimization on Aptitude Area Score, Selection and Classification
OPTFLS	Optimal Assignment, Full Least Squares Prediction
OPTPRFCL	Optimization on Single Composite Predicted Performance, Classification Only
OPTPRFSC	Optimization on Single Composite Predicted Performance, Selection and Classification

P	Form Perception
PAE	Potential Allocation Efficiency
PACE	Professional and Administrative Career Examination
PAT	Programmer Aptitude Test
PC	Paragraph Comprehension
PCE	Potential Classification Efficiency
PDI	Point Distance Index
PSE	Potential Selection Efficiency
PUE	Personnel Utilization Efficiency
Q	Clerical Perception
RDO	Radial Drill Operator
S	Spatial Aptitude
SC	Surveillance/Communications
SDy	Standard Deviation of Performance in Dollar Terms
SIMPO	Simulation of Personnel Operations
SQT	Skill Qualification Test
SR	Selection Ratio
SRQ	Perceptual Ability
ST	Skilled Technical
TC	Tank Commander
u	Unique Component
USAREC	U.S. Army Recruiting Command
V	Verbal Aptitude
VE	General Verbal Ability
WK	Word Knowledge

SUMMARY

The central research question raised in this report is: Can the Army's personnel classification system be improved substantially and, if so, how? The question bears on whether the Army's aptitude area composites are presently adequate and whether the ASVAB tests contain sufficient potential classification efficiency to permit the selection of an adequate set of composites. If the ASVAB were proven inadequate, consideration would need to be given to the development of a new battery with more classification efficient tests or the use of a general cognitive ability composite in place of the existing aptitude composites.

A. CLASSIFICATION EFFICIENCY

Over the last two decades, the tests and test composites (aptitude areas) have been selected to maximize predictive validity with little attention given to improving classification efficiency. We define the classification efficiency of a set of test composites in terms of the gain in mean predicted performance (MPP) score under optimal assignment conditions over that obtainable using random assignment. Classification efficiency depends on allocation efficiency that capitalizes on differential validity and/or hierarchical classification efficiency that capitalizes on heterogeneous validities and/or values assigned to jobs.

The least squares regression weights (LSEs) or full least squares estimates (FLS) applied to all tests forming each test composite of the ASVAB maximize utility when used in either selection or classification. LSEs provide the means of maximizing average validities across jobs and of maximizing potential allocation efficiency (PAE).

The Army's unit-weighted, three-test aptitude composites were standardized to have equal means and variances and are not weighted by either validity or job values. Thus, they do not maximize validities, PAE or hierarchical classification. In contrast, the use of FLS composites would provide a maximum capitalization on hierarchical layering and provide an assured increase in allocation efficiency.

B. THE UTILITY OF ASVAB AND OPERATIONAL IMPLICATIONS

In fiscal year 1987, 315,000 enlistees entered the All Volunteer Force; of these, 130,000, or 41 percent, were recruited into the Army. The services rely heavily on aptitude information (ASVAB), since most recruits have little or no work experience. The services' selection and assignment systems are dependent on an interrelated set of complex factors including policies, goals, recruiting resources, recruiter incentives, formal and informal enlistment standards, the willingness of young people to enlist, and the efficiency of the job assignment system in person-job matching.

The Army's computer-based system Enlisted Personnel Allocation System (EPAS) is being designed to improve the job assignment process by aggregating job demands and applicant forecasts using optimization techniques. Individual assignments are made in real time using a job payoff optimizing technique. Job vacancies and applicant forecasts are updated as applicants receive training seats, and the optimization model is periodically recomputed to adjust assignment guidelines. The optimization model considers 36 job clusters that are similar in terms of performance characteristics and applicant clusters in terms of gender, education, AFQT and aptitude area scores.

We conducted an empirical analysis of productivity gains attributable to simultaneous changes in job entry standards (minimum cutting scores), assignment policies and assignment procedures to provide decisionmakers with realistic information in making rational choices for allocating scarce resources among alternative strategies. Taken together, the need for realism and the need to consider opportunity costs imply that, in order for this utility analysis to be useful, it must be context-specific and credible.

A total of thirty-three different policies were analyzed. Eleven different job assignment policies and procedures were first simulated under 1984 enlistment entry standards, then under the assumption that those standards were raised by five standard score points for all Army jobs, and finally under the assumption of a ten-point across-the-board increase in standards. All thirty-three policies were simulated using the same random sample of 4,280 accessions from 1984 Army enlistments. In addition, to verify the stability of both performance predictions and cost-benefit estimates, nine of the policies were simulated using two different "synthetic" applicant pools.

The use of the full least squares (FLS) assignment strategy permits a shift away from making assignment on the basis of a three-test suboptimally weighted composite. At the same time, a shift also could be made in the objective function being optimized in the

utility analysis. Rather than optimizing aptitude area composites in making assignments, the objective function that could be optimized is mean predicted performance (MPP), the major goal of both the selection and classification processes. In brief, the assignment strategy goal shifts to optimize MPP rather than AA composites.

Using the optimal FLS prediction equations as the assignment strategy raises MPP 0.340 standard deviation units over random selection and 0.143 standard deviation units over the current selection and assignment process; thus, the FLS assignment policy provides 1.7 times the increment of the current policy over random selection and assignment in predicting MPP, using an efficient allocation procedure.

As expected, increasing selectivity (i.e., raising job standards by 5 or 10 points) increases MPP within each of the sets of policies. The use of the FLS assignment policy increases MPP from 0.340 to 0.386 over random selection and assignment with a 5-point increase in standards and to 0.405 with a ten-point increase.

However, simply to know the impact of a policy on performance is not sufficient. Increasing the job standards involves increasing the applicant pool, which, in turn, increases recruiting costs. Therefore, performance gains are evaluated via a benefit-cost model using utility analysis and an alternative economic analysis based on "opportunity costs."

When the optimal full regression equation (FLS) or full LSEs assignment strategy is used, the productivity gain for the first tour of duty is $414 million per year over random selection and assignment. This figure reflects a 73 percent gain in MPP (the benefits component of utility) for FLS over the current assignment system.

The alternative economic opportunity cost model estimates the costs of achieving equivalent levels of performance by increasing the number of high-quality accessions as measured by AFQT scores. Additional recruiting costs are incurred by attracting more high-quality performers that would match the performance of recruits assigned efficiently. For example, the Army could achieve an MPP increase of 0.340 standard deviation units over random by employing the FLS assignment strategy and the selection-ratio employed in 1984, or it could raise the selection ratio (or enlistment standards) to achieve the 0.340 under the current assignment strategy. If the latter approach were to be used, $640 million per year would be needed in recruiting a force comprised of more higher quality enlistees (as compared to the productivity gain of $414 million using an FLS strategy). Thus, an organization that is willing to pay the opportunity costs of $640 million per year in just

recruiting costs to achieve a performance level 0.340 higher than random must believe that the value of productivity, at a minimum, is worth that expenditure of funds.

How much more selective the Army should be must take into account the assumptions made about the recruiting strategies. An increase of five or ten points in all job standards produce increases in the net value of job performance under one set of rational and effective recruiting strategies. However, for another recruiting strategy, making severe assumptions about costs, the raising of job standards ceases to be an attractive alternative. We conclude that it is highly beneficial to increase enlisted standards for the current operational system provided this is not done through simply increasing the proportion of high quality recruits.

Although the present analysis was confined to the Army's selection and classification system, we believe that the results may generalize to the other military services since ASVAB validities and assignment policies and procedures are comparable among the services. If the productivity gains found in the present Army analysis were to be extended beyond the Army's accession of 41 percent of total recruits to include all services, gains for the first tour of duty, attributable to an optimally efficient, but realistic selection and assignment system (i.e., the "constrained" FLS strategy) would reach about $1.0 billion per year over random procedures compared to $494 million for the current system using present job standards.

Our simulation has several limitations that are detailed in the report such as the use of the 40 percent of salary rule-of-thumb. Taken together, these shortcomings most likely resulted in a considerable underestimate of productivity gains attributable to the use of an FLS assignment policy and higher minimum job standards.

We suggest changes that could be made in military operational classification systems that are based solely on our simulation results. The changes depend entirely on better utilization of information contained in the present ASVAB. Only technical changes in assignment policy and procedures are needed to obtain the productivity gains estimated.

Specifically, we suggest the use of: (1) mean predicted performance (i.e., FLS composites) as the objective function, rather than aptitude area composite scores; (2) the least squares prediction equations as the assignment variables rather than equally weighted and reduced numbers of tests of aptitude area composites; (3) an efficient LP allocation algorithm (e.g., EPAS); and (4) raised job standard cutting scores of five standard score points until an optimal allocation is used operationally. The assumption is made that the

preponderance of recruits could be persuaded to accept the jobs in which they can perform best. Such changes appear to be implementable in the near term, given that the assumptions and estimates made in our study hold in the specific decision-context of each service.

However, we do not recommend the implementation of changes based only on simulation results until these changes are considered together with a broader set of empirical findings and psychometric principles detailed in this report.

C. NEW CLASSIFICATION RESEARCH ISSUES

We are following up our simulation study with three research efforts now in progress and one to be initiated shortly. These research efforts aim at increasing the potential classification efficiency of a battery employing basic psychometric principles.

One study employs efficient test selection techniques in a model sampling experiment to maximize potential classification efficiency in the joint predictor-criterion space of Project A. The second model sampling study attempts to utilize differential classification theory applied to the multidimensional structure of the joint-predictor-criterion space to develop optimal ASVAB factor score composites for use in recruit counseling and in record keeping. The third study uses model sampling to compare an ASVAB single stage selection/classification process with the traditional two stage process. Although a simultaneous selection-classification process, multidimensional screening (MDS), holds great promise of utility gains, no empirical evaluations have been reported. The fourth study will determine the upper bounds of gains obtainable from efficiently shredding selected Army job families into a larger number of sub-families.

Underlying these studies is our belief that potential classification efficiency can be improved. The validity generalization movement has provided a great service in pointing out the difficulty of the task. However, it is inappropriate to suggest that the joint predictor-criterion space is inherently unidimensional in nature until a concerted, technically correct effort is expended with the goal of maximizing PCE in both the development and selection of measures for inclusion in the experimental pool. Batteries developed to maximize PSE and validated against limited unidimensional job criteria are not the appropriate reference point for determining the feasibility of an effective classification process. We believe that there is a strong potential for several additional dimensions in the joint predictor-criterion space. Their existence can only be confirmed with the same

concern and care used to identify the existence of general mental ability, clerical speed and psychomotor ability in the joint GATB-criterion space.

D. PROPOSED CHANGES IN THE OPERATIONAL USE OF ASVAB

In the final chapter, the proposed changes for operational use of ASVAB include:

(1) Use FLS composites, that is predicted performance, to provide the maximum amount of PCE obtainable from the present ASVAB; use classification-efficient tests in a modified ASVAB identified by procedures that maximize PCE.

(2) Use job values (as specified by further research) to weight predictor composites, assuming policymakers are willing to explicitly consider values.

(3) Raise minimum cutting scores an average of five standard score units until an optimal allocation algorithm that maximizes predicted performance is employed operationally.

(4) Substitute FLS composites as the measure of quality in place of AFQT to achieve quality goals and in forecasting personnel requirements for systems under development.

(5) Use a generalized FLS composite for applicant screening in place of AFQT to maximize predictive validity of selection (prior to the installation of the two-tiered system).

(6) Shred job families into sub-families and their associated FLS composites to increase the PCE of the present ASVAB.

(7) Install a two-tiered operational system, using FLS composites (transparent to operational personnel) for actual job assignments and factor score composites for record keeping and recruit counseling.

(8) Use a person-by-person optimal assignment algorithm and flexible cutting scores in EPAS to maximize MPP.

(9) Install an integrated two-stage multidimensional screening (MDS) system to make both selection and assignment decisions simultaneously.

While we are confident that these proposed changes could yield productivity gains exceeding 200 percent, we suggest further research and management analysis to determine precise estimates of gains and specification of operating procedures. We show the sequence of implementing the changes over a five year period.

The precise magnitude of dollar savings is not as important as are the relative differences in mean predicted performance among alternative strategies. Our simulation results show that improvement of less than two tenths of a standard deviation unit may

result in gains to the Army of more than $260 million per year. Our proposals are equally applicable to all services and we anticipate comparable gains, subject to confirmation.

The central issue we examined in our analysis concerned the question of the adequacy of the Army's present aptitude area composites and whether the ASVAB tests contain sufficient PCE to permit the selection of an adequate set of composites.

Our simulation results show that the present Army aptitude area composites are of limited value, but there is considerable classification efficiency potentially obtainable from the present ASVAB if used in accordance with differential assignment principles. With further changes in the test content of the ASVAB and with use of classification efficient procedures, we are confident of even greater improvement in potential selection and classification efficiency.

OVERVIEW: THE ECONOMIC BENEFITS OF PREDICTING JOB PERFORMANCE

A. PURPOSE

The central research question raised in this report is: Can the Army's personnel classification system be improved substantially and, if so, how? The question bears on whether the Army's aptitude area composites are presently adequate and whether the ASVAB tests contain sufficient potential classification efficiency (PCE) to permit the selection of an adequate set of composites. If the ASVAB were proven inadequate the alternative would be either the development of a new battery comprised of more classification-efficient tests or one substituting a general cognitive ability for the ASVAB.

Our study has four major objectives: first, to measure the potential gains in Army enlisted soldier performance in each of the Army's nine job families that can be achieved through simultaneous changes in job entry standards (cut scores) and allocation procedures; second, to obtain realistic estimates of the costs and benefits of these performance gains in dollar terms; third, to place these estimates on an evaluative continuum (anchored at one end by the performance levels that would be obtained if the entire process were purely random, and at the other end by the performance levels that would occur if every selected applicant were placed in the job yielding the highest expected performance); and fourth, to make recommendations for improving the current military operational selection and classification system based on psychometric theory, empirical results of previous studies and the findings of our simulation.

Our purpose here is to allow a variety of policies, varying in terms of practical feasibility as well as cost, to be compared to each other in relative terms. The most fundamental requirement for such an effort is that it provide decisionmakers with realistic information that can be used to make rational choices with respect to the allocation of scarce resources among alternative strategies for improving organizational productivity.

B. CLASSIFICATION EFFICIENCY

Over a period of years, the content of the tests comprising the ASVAB and the test composites (aptitude areas) has been selected to maximize predictive validity with little attention given to improving classification efficiency. Both psychometric principles and empirical results show the emphasis on predictive validity and on operational simplicity (a carry-over of a precomputer age) to be fundamentally erroneous.

We define the classification efficiency of a set of test composites in terms of the gain in the mean predicted performance (MPP) score under optimal assignment conditions over that obtainable using random assignment. The potential classification efficiency of a battery is defined as the gain in the MPP score resulting from optimal assignment obtainable using full least squares (FLS) composites as both assignment and evaluation variables.

Classification efficiency depends upon classification processes, allocation efficiency and hierarchical classification efficiency. The allocation process capitalizes on differential validity; all classification effects are explainable as either allocation or hierarchical classification resulting from the disparate means and variances of criterion variables. When heterogeneous validities and/or values are assigned to jobs and are also reflected in the *prediction variables used in the assignment process, hierarchical layering effects result.*

The least squares regression weights (LSEs) applied to all tests forming each test composite of the ASVAB maximize utility when used in either selection or classification. Such composites will not only provide the means of maximizing average validities across jobs, but will also maximize potential classification allocation efficiency (PCE). The validities of the composites are the multiple correlation coefficients between the composites and each job criterion measure. If the composites use a reduced number of tests or are not LSEs, the best composites for selection are not necessarily the best for classification. LSEs maximize both validity and the PCE obtainable from the battery.

A difference among mean benefit scores across jobs can result from either differences in validities or from the differences in value accorded to jobs (both differences may exist in the same situation). To capitalize on differences in validities (i.e., hierarchical classification), the most effective composites will be the least squares predictors, the actual predicted benefits.

The current Army aptitude area (AA) composite predictors, even using an optimal assignment algorithm, would not elicit the hierarchical layering effect since the composites

were standardized to have equal means and variances and are not weighted by either validity or job values. Therefore the Army's current use of AA composites as assignment variables does not maximize validities or PAE, and has zero hierarchical classification efficiency.

In brief, the current operational assignment system, which attempts to maximize AA composite scores as the objective function, needlessly reduces both validity and classification efficiency compared with an assignment strategy that uses full regression equations to maximize MPP as the objective function.

If AA composites are converted to standard scores and multiplied by their validity coefficients, the composites could contribute to hierarchical classification. The use of FLS composites (not standardized to provide equal means and variances across composites) would provide a maximum capitalization on hierarchical layering as well as an assured increase in allocation efficiency.

The Army's problems with having an ineffective set of assignment composites (AA) is complicated by the need to change policy if the benefits of the best replacements, FLS composites, are to be realized. Classification efficiency also could be improved by making job families more homogeneous, raising minimum cutting scores, and by providing a greater role for optimal assignment algorithms.

A number of previous research results are reviewed pertaining to the adequacy of the AA composites and whether the ASVAB contains sufficient PCE to permit the selection of an adequate set of composites. We conclude that the Army AA composites, as currently used, are of questionable value, but that considerable classification efficiency is potentially obtainable from the existing ASVAB if it is used in accordance with differential assignment theory. The theory focuses on classification efficiency as measured by MPP using a specified assignment procedure. Any gain or loss in predictive validity is relegated by the underlying mathematics (a result, not an assumption) to what in many cases plays a minor role in achieving classification efficiency improvements.

C. SIMULATING SELECTION AND ASSIGNMENT POLICIES

Previous applications of utility analysis for the purposes of benefit-cost analysis suffer from several limitations. Virtually all examples of the utility of testing deal with very simple application models. Testing is usually applied only to selection for a single job. The employer can either pick candidates from the top down from a batch of applicants, or a

single standard can be enforced for an extended period of time. The supply of applicants available to an employer is given. There are no major psychometric problems of selected applicants turning down the employer, negotiation over employment conditions, long term retention or filling more than one job simultaneously. (In connection with job offers, Murphy (1986) discussed the effect of rejected offers on selection utility and Schmidt et al. (1979) discussed means of an adjustment of the normal curve to allow for rejections.)

This report deals with the application of testing to the U.S. Army. The procedures it employs for selection, classification, and allocation are considerably more complex than the non-military examples that exist in the literature. The Army is the nation's largest single employer: each year 130,000 new recruits are selected for 258 different entry-level jobs. The Army selects and assigns recruits to these different jobs in a two-stage process. Chapter 2 describes the manner in which this is done and the various organizational goals and constraints that affect the enlistment process and hence the use of testing. The need to fill a variety of different training classes from a heterogeneous pool of applicants presents a complex management problem for the Army.

Despite the immense management problem faced by the Army, progress is being made in incorporating more use of selection and classification measures into personnel assignment. A new system that makes improved use of information on applicant attributes, forecasts of the composition of the applicant pool, and explicit allocation objectives, currently under development, is described in Chapter 2. This system, the Enlisted Personnel Allocation System (EPAS), can also be readily adapted to use new predicted performance composite information evaluated in this report.

EPAS, an operational personnel management system that attempts to optimize performance goals along with meeting manpower policies, provides a realistic way to improve the use of test information. However, a key question that management wants answered is what such improvements are worth to the Army. Chapter 3 addresses the measurement of the benefits and costs of selection and classification policies.

Chapter 3 also develops a general model of selection and classification decision-making. This model, based on classical economic production theory, incorporates several key aspects of human resource management that are typically omitted from utility analysis. The relationship of recruiting and training to testing is considered explicitly. In order for an organization such as the Army to increase its selection ratio, it must increase its recruiting costs.

A second feature of the analysis provided in Chapter 3 is the consideration of a variety of alternative policies and procedures. Both theoretical and realistic policies are evaluated. Policies range from random selection through operational assignment systems and maximum classification schemes. Also, alternative job standards are explored, while the combination of scenarios are explored simultaneously.

In addition to considering a wide range of scenarios, the simulations measure the impact of different classification strategies using several populations. Aptitude area composites, full least squares estimates and several other types of composite are used as alternative assignment policies. Three different samples are assigned and allocated empirically by means of EPAS to investigate the effect of population variability: a sample of 1984 recruit accessions, a synthetic sample representative of the youth population as a whole, and a synthetic sample that resembles current selection standards.

Two approaches are used in Chapter 3 to convert predicted performance changes into benefits and costs. First, the traditional ut[illegible] is broadened to account for recruiting and training effects. Second, an economic opportunity cost model is applied to the alternative policies considered. This model estimates the additional resources that would be required to achieve a given performance level under current policies. For example, an alternative to improving the assignment system would be to maintain the current assignment system, but allocate more resources to recruiting. Since considerable information is available on the cost of recruiting, it is possible to infer a value for such improvements through recruiting costs.

The results of Chapter 3 are innovative in a number of ways. The application of testing procedures produces impressive benefits to the Army in terms of increased performance and lower attrition, as one would expect from traditional utility analysis. However, the payoffs from operational policies are likely to be considerably less than suggested by theoretical utility analysis when costs are fully considered. Nevertheless, the more realistic evaluations similar to the ones described here were found to be convincing to management, since they deal with many of the operational issues that must be addressed in the implementation of program changes.

Another innovative aspect of the research in Chapter 3 is the comparison of the results of changing either job standards or assignment procedures, or both. The benefits of increased job standards are sensitive to assumptions made concerning recruiting costs. At some point it becomes more expensive to increase standards than gains to productivity

warrant. While the gains from classification may not appear to be as large as gains from selection, they are more robust in terms of net benefits to the Army. That is, there is very little operational cost involved in achieving allocation efficiency.

In brief, the simulation model developed here, together with its accompanying expansion of benefit-cost analysis, provides a number of significant advantages. First, we have greatly expanded the capacity to simulate alternative personnel management policies. Alternative selection, classification, and assignment policies can be simulated in considerably more detail than was possible before. The outcome of these policies can be examined not only against aggregate outcome measures, such as predicted performance and attrition, but can be analyzed in detail by job family or category of recruit. Furthermore, alternative scenarios with different requirements and applicant pools can be evaluated readily.

The approaches to evaluating outcomes has been similarly expanded. We provide two alternative benefit-cost methodologies: one output-oriented, based upon psychological utility theory, and an alternative input-oriented opportunity cost theory. Both methods can readily be adapted to new assumptions of training and recruiting costs or SD_y.

D. UTILITY OF THE ASVAB

Our simulation results show, using the lowest recruiting cost assumption, that the net present value increases for all policies that improve assignment. That is, benefits are higher when recruiting and training costs are taken into account. Using a very conservative estimate for the gains of EPAS (aptitude area score optimization, rather than predicted performance), it is possible to increase productivity by 56 million dollars annually under the present assignment system.

The optimal full least squares solution (FLS) demonstrated by far the greatest potential benefits. Over 260 million dollars annually to the Army in performance gains could be achieved under this policy.

Under the "medium" recruiting cost estimate, the performance gains that would be produced under current assignment policy increases to $22 million by raising the enlistment standard by 5 points and $16 million by raising the standard by 10 points. Other assignment strategies also show a significant net benefit to increased selection under this cost estimate. For example, under full least squares assignment (OPTFLS) the net present value of productivity gains would be worth about $278 million annually to the Army. Very

similar results of the expected performance gains are produced under the "low" recruiting cost option.

However, under the "high" recruiting cost option, the most conservative strategy employed, raising minimum job standards ceases to be an attractive alternative. The high recruiting cost option assumes a recruiting strategy that meets increased standards by simply increasing the number of high quality recruits (I-IIIA), rather than screening more recruits over the same quality range.

The recruiting cost assumption behind the low and medium estimates is that the need for additional qualified recruits is largely met by screening additional IIIB and IV candidates. Thus, these two policy options show it is beneficial to increase standards, even under current assignment procedures. The importance of recruiting cost assumptions and policies becomes evident when one examines different job standards. If job standards can be met primarily through screening a larger pool of applicants, it is cost-effective to raise standards, either under the present allocation system or an improved system such as EPAS or FLS. However, if standards must be met through increasing the proportion of more highly qualified applicants, then it ceases to become an attractive alternative.

Using the "opportunity cost" approach, in place of the net present value (dollar value of a standard deviation in performance), we ask, "What would it cost to achieve the levels of performance produced under each evaluated policy if the mechanism used to achieve those gains were simply to increase the numbers of high quality recruits and assign them using the current system?"

Using recruiting opportunity costs as a measure of the benefits produces results that are generally higher than the net present value approach. The largest difference is for the opportunity cost of FLS assignment. Such a policy would require 81 percent I-IIIA recruits under current enlistment standards, and an Army comprised nearly entirely of high-quality personnel under an enlistment standard raised by 10 points. The annual benefits of optimal FLS assignment increase dramatically for such a high quality force, since recruiting costs increase at a quadratic rate. Opportunity costs of such a policy are nearly $640 million annually under current standards, and over $993 million annually under the most restrictive enlistment standards. It should be noted, then, that the net present value utility gain of $414 million attributable to our efficient selection and assignment policy (FLS), under current standards, appears conservative in contrast to just the opportunity cost of $640 million of recruiting equivalent levels of performance.

Although the present simulation was confined to the Army's selection and classification system, we believe that the results should generalize to all the military services since ASVAB validities and assignment policies and procedures are comparable across the services. For example, if the productivity gains found in the present simulation were extended beyond the Army's accession of 41 percent of recruits to all military services, gains attributable to an optimal selection and classification system would be about 1.011 billion per year over random selection and classification. In contrast, the gain attributable to the current system is $494 million.

The simulations made thus far have produced a number of important findings. First, assignment policy can be improved greatly using EPAS. Second, it is highly beneficial to increase enlistment standards provided this is not done simply through increasing the proportion of high quality recruits. Third, by the use of FLS composites to predict performance differentially, it may be possible to more than double the benefits from assignment.

There is likely to be much greater classification efficiency and payoff from psychometric research that improves differential validity and employs differential assignment technology than through any other approach, such as predictive validity. Once a system such as EPAS is implemented, it is likely that there will be substantial payoff from using FLS composites of the existing ASVAB and still greater improvement if we identify and incorporate new classification-efficient tests and efficient composites into the Battery.

Although our study is aimed at incorporating accuracy in parameter estimates and realism in assumptions, it has several limitations that are addressed, including: limiting the number of jobs that were sampled and their "representativeness"; using lower bound SD_y estimates equal to 40 percent of salary; using only one component (technical proficiency) from among five distinct components; using FLS weights based on the validities of aptitude area composites rather than on the validities of the ten subtests of ASVAB; failing to achieve the full potential of hierarchical layering effects by valuing jobs equally; failing to subject the parameters used in the analysis to a risk analysis (e.g., the sensitivity analysis of using job standards did not identify a precise estimate of the optimal job standard and recruiting strategy); underestimating of the prediction of attrition by the EPAS system. Taken together, these shortcomings most likely resulted in a considerable underestimate of productivity gains attributable to the use of an FLS assignment policy and higher minimum job standards.

In one instance we used parameter values that would most likely result in overestimates of gains. The same weights were used for the identical set of assignment and evaluation variables in measuring mean predicted performance. Thus correlated sampling error was incorporated in the measure, although we believe the effect was more than equaled by that of the conservative estimates guaranteed to provide underestimates of gains.

E. OPERATIONAL IMPLICATIONS OF THE SIMULATION

We now address the changes that could be made in the military services' operational classification systems, based solely on our simulation results. Only technical changes in assignment policy and procedures are necessary to obtain productivity gains of the levels estimated in the present study. The changes call for the best use of all information contained in the present ASVAB along with a simultaneous increase in job standard minimum cut scores.

Specifically, we suggest four changes that appear to be implementable in the near term, given that assumptions and estimates made in our study hold in the specific decision-context of each service: the use of mean predicted performance as the evaluation function; the use of full least squares prediction composite for each job family; the use of an efficient computer-based algorithm to allocate personnel using predicted performance as the objective function; and the raising of job standard minimum cut scores by five standard score units until an optimal assignment system is used.

The assumption is made that the preponderance of recruits can be persuaded to accept the jobs in which they can perform best or nearly best.

The major implementation effort required for operationalizing these recommendations involves the development of an efficient linear program (LP) computer-based algorithm for assigning individuals. The developmental work for such an assignment system is being accomplished in EPAS. Two minor modifications of EPAS would be required for incorporating an LSEs system providing the productivity gain estimated in this study: frequent updating of the allocation plan (e.g., once every two weeks) and the addition of a "column constant" to each recruit's LSE scores for each job family to meet policy constraints.

We are not, however, recommending the operational implementation of these possible changes based only on simulation results until they are considered together with a broader set of empirical findings and psychometric theory described below.

F. NEW RESEARCH ISSUES ON CLASSIFICATION EFFICIENCY

We are following up our simulation study with three research efforts in progress and one to be initiated shortly. These research efforts aim at increasing the potential classification of a battery employing basic psychometric principles.

A set of test composites can provide no more PCE for a prescribed set of job families than was provided in the test selection process that created the operational test battery. PCE can only be increased for a fixed operational battery by efficiently increasing the number of job families with their associated predictor composites. Conversely, if the number of job families is specified and the test battery is not fixed, PCE can be improved by efficiently selecting tests for use in a new or modified battery. Applying such principles suggests a number of possible changes.

The FLS composites already provide the maximum amount of PCE for a fixed battery and specified set of jobs or job families. However, improvement in PCE can be accomplished by selecting predictors that experts believe have a high degree of differential validity (as contrasted with predictive validity) for inclusion in the experimental test pool, and by test selection using indices that measure PCE to identify the operational battery with the best PCE.

Given that a small number of FLS composites are being used to assign personnel to the same number of efficiently determined job families, a worthwhile improvement in MPP can be obtained by a major increase in the number of job families. An increase in the number of composites and associated families to somewhere between 20 and 40 would most likely provide the maximum efficiency for Army jobs.

The use of numerous test composites would require the Army to record many scores on official records. One way to use many assignment composites is to install a two-tiered system in which the large number of FLS composites are used to make recommendations regarding assignment, while a much smaller number of factor scores are used for counseling.

The largest increase in MPP will undoubtedly come from the use of FLS composites for both selection and classification in the distinct two-stage operational process now employed. Further worthwhile improvements may result from the use of a single process that enables selection and classification decisions to be made simultaneously.

The four promising operational changes outlined below are to be investigated in a series of studies.

(1) Replace or augment existing ASVAB tests with new predictors selected from Project A experimental variables, using a test selection index which maximizes PCE rather than predictive validity.

(2) Determine the PCE provided by several sets of factor scores (composites yielding factor scores) compared to Army AA's and FLS composites; the factors on which scores are based will be obtained using an approach which maximizes PCE.

(3) Determine gains in MPP obtained from the use of MDS by comparing the traditional two-stage strategy with two simultaneous selection and classification strategies.

(4) Determine the upper bounds of gains obtainable from shredding selected Army job families into sub-families, then estimate gains in MPP obtainable from increasing the number of job families using an optimal clustering algorithm that maximizes PCE.

The studies outlined above are described in detail in the Appendices of Chapter 5. The designs employed illustrate some of the features of model sampling experiments. Other studies for future effort include: influencing applicants in their decisions to accept those jobs they can perform best, developing new test measures that increase PCE, developing new utility measures that consider job values and a broader array of criterion measures, and evaluating new procedures in the field context.

G. PROPOSED CHANGES IN THE OPERATIONAL USE OF ASVAB

In Chapter 6 we recommend changes in the operational use of the ASVAB. On the basis of our simulation findings, prior research results and psychometric principles, we conclude that very large productivity gains can be achieved principally by changing the policies and procedures that govern the operational selection and assignment system. We propose a sequence of changes that are implementable over a period of several years, provided our assumptions and estimates are confirmed in the specific decision-context of each service.

The proposed changes are:

(1) Allocation Efficiency

- Use FLS composites in standard score form that resemble AA composites for an initial period of time to capitalize on differential validity to improve PAE.

- Use FLS composites converted to predicted performance after the initial period to provide maximum amount of PCE obtainable from the present ASVAB.

(2) Hierarchical Classification

- Use job values across different jobs and/or values for different performance levels within a job to weight predictor composites, assuming policymakers are willing to consider values explicitly.

(3) Raise Minimum Job Standard Cutting Scores

- Use cutting scores raised an average of five standard score units, resulting in productivity gains of about 21 percent over current procedures not employing an optimal allocation algorithm.

(4) Substitute FLS Composites in Place of AFQT as Quality Measure

- Use FLS composites as quality goal measures in place of AFQT to distribute quality, to raise predictive validity in a job family and PAE across job families.

(5) Use a Generalized FLS Composite for Recruit Selection

- Using all the predictors of the ASVAB in a generalized FLS composite, rather than the AFQT, would maximize the predictive validity of recruit selection.

(6) Use Additional Job Families

- Increasing the number of efficiently determined job families and associated FLS composites would result in large productivity increases through increases in PCE.

(7) Develop and Implement a Two-Tiered Assignment System

- Use the FLS composites for actual assignment to 20-40 job families, but use only sets of factor score composites as the visible system for record keeping and recruit counseling.

(8) Use Improved Person-Job Matching Algorithms

- Use both predicted performance and attrition as the variables to be optimized in EPAS assignments rather than aptitude areas
- Use person-by-person assignment procedures to maximize MPP
- Use flexible cutting scores in making assignments to maximize the mean assignment variables

(9) Use an integrated multidimensional screening (MDS) system

- Rather than selecting applicants in one stage and assigning recruits in a distinct second stage, make both selection and assignment decisions simultaneously for further productivity gains.

Although we are confident that these proposed changes could provide immediate benefit if implemented today, we suggest further research and management analysis to determine more precise estimates of productivity gain and how to make the most efficient applications of the proposed new procedures. We show the sequence of implementing the changes over a five year period.

Our ball park estimate of productivity gains attributable to improved PCE procedures may approach 200 percent; productivity gains attributable to improved selection and MDS may be between 15-25 percent.

The precise amount of dollar savings is not as important as are the relative differences in mean predicted performance among alternative strategies. We know from our simulation results that improvements of one- or two-tenths of a standard deviation of MPP may result in a very large gain. For example, in the Army a 0.143 gain in MPP results in more than a $260 million gain each year for FLS composites compared to current AA composites.

Although our simulation was accomplished using Army data and our other analyses also focused on data in the Army context, we feel the proposed changes are equally applicable to all services and we expect comparable gains, subject to confirmation.

Our analysis shows that the current Army aptitude composites are of limited value, but we also show that considerable classification efficiency is potentially obtainable from the present ASVAB if the battery is used in accordance with classification-efficient procedures. The ASVAB would have possessed even more PCE if its development had not been based largely on a search for increasing the validity of specific aptitude tests rather than on a search for increasing MPP. The proposed changes we suggest offer almost certain promise of large improvements in selection and classification efficiency.

CHAPTER 1. THE PSYCHOMETRIC BASIS OF PERSONNEL CLASSIFICATION

The central research question raised in this report is: "Can the Army personnel classification system be improved substantially, and if so, how?" First, in this chapter, we contrast the deficiencies of the Army aptitude area (AA) test composites currently used to classify and distribute personnel with the potential effectiveness available from current and hopefully improved future predictor variables. We then provide a brief survey of psychometric principles that apply to any effort directed at the evaluation of, and/or improvement in, the classification efficiency of test composites.

Next, we discuss two major challenges to the concept of using a set of differentially valid test composites coupled with an optimal assignment process for matching personnel to jobs. The first of these challenges is posed by those who maintain that a single measure of general cognitive aptitude is sufficient to explain the predictability of job performance criteria. The second challenge is from those who doubt that it is possible to transform performance measures into a metric that adequately represents the benefits from improving performance across different jobs or that can be used to trade off costs against the benefits of improved performance.

Finally, we consider alternative simulation approaches to answering our central research question. We select an approach which utilizes scores from an available data base for use in the simulation--instead of a model sampling technique in which synthetic test scores are generated for use in a simulation of the classification system.

A. THE OPERATIONAL PROBLEM

The operationally extant Army personnel classification and person-job matching system in 1988 utilized a set of nine aptitude area test composites corresponding to nine job families that evolved from two decades of research emphasis on enhancing predictive validity. The content of both test composites and the operational test battery, Armed Services Vocational Aptitude Battery (ASVAB), has been selected to maximize predictive validity--with little or no attention paid to improving the classification efficiency of the total set of test composites in a multi-job, optimal assignment situation. Traditionally, the

number of tests per composite has been kept small and the weights restricted to unity--or, at most, to two or three--in order to simplify the operational use of the composites. This emphasis on predictive validity and its operational simplicity (required in a precomputer age) can be shown to be either outdated or fundamentally erroneous with respect to both empirical results and psychometric theory.

Test composites with a moderately high degree of classification efficiency would show greater validity for the associated job family than they would show, on the average, for the other job families. McLaughlin, Rossmeissl, Wise, Brandt, and Wang (1984) provide a table of adjusted validities of the nine Army AA composites for each of the nine job families reflecting the combined results of recent studies using either SQT or training performance as the criterion variable (p. 22). Only two of the nine AA composites indicate an acceptable level of classification efficiency: Clerical/Administration (CL) and Skilled Technical (ST). They showed a difference between the validity for the associated job family and the average validity across all job families of +.08 and +.10 respectively. This difference for the other seven AA composites was −.05, −.01, −.01, −.01, −.01, .00, and +.01. Only the ST composite had its highest validity for the corresponding job family.

It appears that the existing AA composites are heavily saturated with a predictive component of general cognitive ability that is generally valid across all job families. At some level of saturation with this generally valid measure, the use of a single measure of general cognitive ability in lieu of the AA composites would seem appropriate. However, the use of a single measure such as the Army General Classification Test (AGCT) that preceded the Army Classification Battery (ACB) and ASVAB, or a deliberately crafted measure of general cognitive ability from the ASVAB, could make a contribution to classification greater than would the use of random assignment only if current Army policy were changed to permit higher quality (i.e., high scoring) personnel to be assigned to the more intellectually demanding jobs.

It is clear that the amount of predictive validity provided by each test composite of a battery is a very poor indicator of classification efficiency provided by a set of composites. *Selection* efficiency is commonly measured in terms of the mean predicted performance (MPP) of those selected for the job. This MPP value can then be converted into a benefits measure that is compatible with a cost measure; the utility of a given selection procedure is often computed in terms of dollars. This approach to determining utility assumes the existence of a common metric which can represent both benefits and costs.

Similarly, one can measure the effects of a *classification* procedure in terms of MPP. A value of MPP across several jobs resulting from a classification procedure, involving both specified composites and assignment algorithms, can also be converted to a utility measure, if one assumes the adequacy of the benefits metric for measuring the value of performance across jobs. Since we are interested in converting classification efficiency into utility, prior research results expressed in terms of MPP are most relevant. The simulation research presented in Chapter 3 first provides MPP values and then converts these values into dollars.

The fact that predicted performance can be substituted for the actual performance measures in the determination of either selection or classification benefits will be elaborated in a later section. It is clear that MPP is equal to the mean performance measure expressed in standard score form multiplied by the validity coefficient of the Full Least Squares (FLS) corresponding to a given job or job family. If the quotas for all jobs were equal and the correlations among predicted performance (PP) scores were also equal, the expected MPP score for a particular job would be directly proportional to the validity of each PP score. The higher the validity, the higher the expected MPP score. Also, these expected MPP scores for each job usually will be higher for jobs which have lower average intercorrelations of its specific PP variable with all other PP variables, and higher for jobs with lower quotas.

It is also true that an FLS composite using weights based on sample estimates will have for each job an expected mean score proportional to the MPP score (based on universe weights) for that job. The expected variances of the FLS composites and the universe PP variances will also be proportional. This close relationship expected between FLS composites and the universe PP measures when the parameters of the assignment and evaluation variables are computed in separate cross samples becomes exact when the parameters of both variables are the universe values.

Even when there is only one measure, or no reliable independence among the composites used for both selection and classification, the MPP score across all jobs can be increased as a result of the assignment process. This can occur under these circumstances when the rank order of MPP scores across jobs closely matches the rank ordering of the mean scores for the composite used to make assignments to each job. This layering of mean scores, with the mean assignment score and the MPP score for the same job generally falling into the same layer of rank ordered means, is referred to as hierarchical layering

(Johnson and Zeidner, 1989). We call personnel classification that achieves this layering effect, "hierarchical classification."

The Army AA composites are standardized so that they have a mean of 100 and a standard deviation of 20 in the youth population. Such equating of expected means and variances across composites assures that the Army classification process cannot capitalize on hierarchical layering that might exist if the AA composite scores were converted, by a change of scale, into predicted performance (PP) scores. We call this kind of classification, which does not rely on hierarchical layering, "allocation" (Johnson and Zeidner, 1989). The capability of a set of variables with equal means and variances to increase MPP scores through the use of an optimal assignment algorithm we call "potential allocation efficiency (PAE)."

If AA composites are converted to standard scores and multiplied by their validity coefficients, the composites thus adjusted are capable of hierarchical classification. FLS composites, unless they are standardized to provide the same equality of means and variances across composites as is present in AA composites, will provide a maximum capitalization on hierarchical layering, as well as a guaranteed increase in allocation efficiency.

The Army's problem with having an ineffective set of assignment composites, the Army aptitude areas, is complicated by the need to change an existing policy if the theoretically best replacements, FLS composites, are to be utilized. A part of the considerable gain in classification efficiency one can expect from substituting FLS composites for the existing AA composites (an expectation forecast by prior results reported in the following section) is a result of utilizing the advantages hierarchical classification often shows over allocation. The implications of such a policy change will be discussed in greater detail in Chapter 4.

Changing to an all-volunteer Army has also contributed to a decline in the effectiveness of the Army classification system. Use of an optimal personnel assignment algorithm gave way to reliance on minimum cutting scores to achieve any benefits obtainable from the use of a set of composites that could not be provided by a single measure. At the same time, the cutting scores for the various jobs or military occupational specialties (MOS), over time, became both lower and more similar. Fewer soldiers are denied their preferred assignment because of failure to meet prerequisites, and the AA composites have less effect on the assignment process.

One alternative approach to relying upon the classification battery to play a significant role in the assignment process is to have a recruiter or job counselor provide the potential recruit a more focused choice of occupation along with information on his or her predicted performance. There is some evidence that a counselor can greatly influence the choice of the potential recruit, i.e., that a counselor can sell the applicant on selecting from a set of jobs for which aptitude has been demonstrated to be high.

There are several ways in which classification efficiency of the present system could be improved: (1) providing a more effective set of the nine operational test composites; (2) using composites with means and variances that reflect the extent that performance in the associated job family is predicted by the composite; (3) shredding out the existing job families in order to make them more homogeneous and readily represented by the associated composite, at the cost of reducing the number of cases on which the validity vector is based; (4) raising minimum cutting scores for jobs which realistically have higher prerequisites (and usually have higher validity); and, (5) changing the recruiting system to provide a greater role for optimal assignment algorithms. Previous research, particularly Sorenson's (1965) and that of McLaughlin et al. (1984) indicate the probable desirability of using FLS composites that combine the first and second ways above. Simulation results could be used to determine the utility of introducing the changes required to effect 1 through 5.

While prior results clearly indicate that FLS composites with regression weights computed on very large samples can be expected to have more PCE than the existing Army AA composites, these reported results are not expressed in terms of utility. Policymakers have been reluctant to disturb the status quo in order to improve a psychometric index, e.g., validity, even though this index is purported to be closely related to PCE. They apparently are equally reluctant to support changes in the operational system to achieve the goal of increasing the obtained MPP scores, even by as much as 100 percent. The benefits of such changes must be expressed in a metric that permits the trading off of costs against benefits in an utility context, if decisionmakers are to be persuaded to support change.

Thus, we concluded in our planning phase that to be effective, a comparison of FLS composites with the existing AA composites must be made in the framework of a utility analysis, using the state-of-the-art knowledge of psychometric principles relating to classification, and the large-scale current data available from Project A. To this end a research team was assembled to design and implement a simulation experiment that would provide the benefits side of a utility analysis. This team was comprised of an economist

skilled in cost analysis and other quantitative and computer skills, an operations research scientist directing the ongoing development of a new Army personnel classification, person-job matching and distribution system, and two I/O psychologists knowledgeable in the psychometrics of classification and in utility measurement. This team reached a consensus on how to approach the simulation which was accomplished by the authors of Chapters 2 and 3.

B. PRIOR RESULTS SPECIFICALLY RELEVANT TO THE PROBLEM

In this section we review research evidence on two related topics: (1) the PCE of the Army AA composites, and (2) the PCE of the most effective composites created from the ASVAB corresponding to the existing nine job families. The first topic bears on whether the Army AA composites are presently adequate, the second on whether the ASVAB tests contain sufficient PCE to permit the selection of an adequate set of composites. If the ASVAB were proven inadequate, the alternative would be the scrapping of the ASVAB in favor of either a general measure or a new battery of more classification-efficient tests.

The optimal test composite for use in either selection or classification is of course the least squares estimate (LSE) of job performance. To be optimal, this "best weighted" composite of tests must include all tests in the battery. Such an LSE measure is referred to as a full least squares (FLS) composite. Sets of test composites that use a subset of the total battery for each composite are not optimal for either selection or classification. The "best" reduced size subset to be used to form a given composite depends on whether it is desired to optimize the effectiveness of the set of composites for selection or for classification. This and other psychometric principles governing classification will be elaborated upon in the following section.

McLaughlin, Rossmeissl, Brant and Wang (1984) compare the classification effectiveness of the Army AA composites, FLS composites, and other composites using an index called M^2. These authors believe their index measures differential validity for non-FLS composites with results that are comparable to the use of Horst's index of differential validity on FLS composites. Horst's index, restricted to the evaluation of the classification effectiveness of a battery when only FLS composites are used, could not be used to measure the effectiveness of the Army AA composites (Horst, 1954).

The Project A study described in McLaughlin et al.'s report provided for the collection of both operational and experimental data on over 60,000 soldiers and 98 jobs

(MOS). Only the existing ASVAB tests were considered in research to determine the advisability of reconstituting the operational AAs and restructuring Army job families. Using test intercorrelations and test validities against training or Skill Qualification Test (SQT) performance criteria for a large number of soldiers assigned to the 98 different jobs, full regression weights for each job were computed. FLS composites could thus be compared with other sets of test composites, including both the Army AA composites and a single measure of general cognitive aptitude (i.e., as if a current version of the AGCT were to be used in place of its successors, the ACB and the ASVAB).

McLaughlin et al. (1984) used an average of the Horst differential efficiency index (H_d), designated by them as H^2, and the creative extension of the concept of H^2, designated as M^2, to measure the potential classification efficiency of the alternative AAs. They proposed the ratio of (M/H) as an estimate of the percentage of total differential validity that could result from optimal use of aptitude areas. This they contrasted to the optimal utilization of the ASVAB (98 FLS composites) to assign soldiers to the 98 jobs using an assignment algorithm that maximizes the predicted performance (PP) of assigned personnel. They refer to this percentage as "relative efficiency," and say that it assesses "the extent to which the composites capture the differential validity possessed by the ASVAB" (p. 49).

As described in Johnson and Zeidner (1989), H_d is the sum of the squared correlation coefficients between two differences associated with each pair of jobs. One of these arrays of differences (the criterion differences) is between either the actual performance measures or the predicted performance measures (the use of either one would yield the same result). The arrays correlated with the criterion difference arrays, the designated predictors of the criterion differences are the differences between the two predictors corresponding to the two criterion variables making up each unit of analysis. Horst prescribed using FLS composites as the predictors in his formulation of H_d. The Project A authors define the "predictors" as least squares estimates (LSEs) based on the two AAs corresponding to each criterion pair.

The computational procedures devised by these authors included several desirable refinements in algorithms used for H^2 and M^2. For example, alternatives were provided for both algorithms in which the number of soldiers assigned to each job is taken into account. Also, the LSEs for performance on each of the 98 Army jobs are adjusted using the ridge equation method to reduce shrinkage of validity of these best-weighted equations in future samples (Draper and VanNostrand, 1979). Appropriately, in the computation of

M^2, the same estimates of performance differences are used across the different batteries (i.e., the different sets of AAs). These added computational features make the comparison of M^2 values more meaningful across sets of AAs than if an approach similar to that used in Horst's (1954) examples had been utilized.

The authors reported the "relative efficiency" of the composite set comprised of 98 LSEs (i.e., one per job in lieu of AAs, and measured in terms of H^2), as 100 percent (by definition). The composite set of the current 9 AAs has a "relative efficiency" of 64 percent and a single AGCT type composite has a relative efficiency of 43 percent, where the more traditional formulae for H^2 and M^2 are used, i.e., job samples are not weighted by their size. Additional results are provided in Table 1.1.

Table 1.1. Differential Validity Indices for Alternative Sets of Test Composites

Composite Sets	Differential Index (*H* or *M*)[a]	
	Traditional Index (unweighted by Job Density)	Index Modified to Reflect Job Density
98 LSEs	0.314	0.214
Current 9 Aptitude Areas	0.202	0.146
Revised 9 Aptitude Areas	0.190	0.142
4 Composite Set	0.160	0.125
3 Composite Set	0.154	0.120
2 Composite Set	0.150	0.125
1 Composite Set	0.136	0.106

Source: Adapted from McLaughlin et al. (1984), pp. 50-51.

NOTE:

a *H* is used for the LSEs and *M* for all other composites; *H* is the square root of the mean value of Horst's index of differential validity, thus $H^2 = (H_d)/m$; *M* lacks a precise relationship to *H* (see text for description of *M*), but McLaughlin et al. appear to believe that *H* and *M* are comparable.

The revised set of 9 AAs, as recommended in the McLaughlin et al. report, show an 18 percent reduction in the gain of M^2 provided by the 9 operational AAs over the single AGCT type composite (again using the unweighted formula). It is noteworthy that the

authors considered such reduction in differential validity an acceptable price to pay for an increase in predictive validity.

We believe M^2 has serious limitations as a measure of differential validity, and H^2 (or H_d) has an unknown relationship to mean predicted performance under the conditions presented by the Project A data. However, this study provides a strong indication that the existing 9 AA composites provide approximately half as much improvement in classification efficiency, as compared to the use of FLS composites, over the use of a single measure in a hierarchical classification situation. It is surprising that replacement of the AA composites with FLS composites is not a recommendation they make on the basis of their study.

A later analysis of Project A data using LISREL (Wise, Campbell, Peterson, 1987) shows that the hypothesis that a single best weighted composite fits all jobs could not be rejected separately using each of the five criterion composite measures other than for technical proficiency--three "will do" criterion components in addition to two "can do" components. The authors state that "For Core Technical Proficiency, however, the common prediction equation model was strongly rejected" (p. 5).

The four Project A criterion components for which a single measure was an adequate predictor of performances across jobs should not be expected to have differential effects across jobs. The "can do" measure was intended to reflect common military skills required of all soldiers; and the "will do" components were attempts to measure personal characteristics we believe to be no more than lightly influenced by the job to which a soldier was assigned.

The Wise et al. study provides interesting research results that indicate promising multidimensionality in the Project A data. We are passing lightly over these results because they used variables not included in the ASVAB and because they did not provide information bearing on the utility of using job-specific, FLS composites.

Most empirical comparisons of the Army AA composites with a single "general" measure and the more job-specific FLS composites have been made on the basis of predictive validity. We will explain further in Section D why such studies lack relevance for classification efficiency; the average predictive validity of composites across jobs is a comparatively minor ingredient of the potential classification efficiency of sets of composites used as assignment variables in an optimal assignment process. Thus, we will skip over the vast literature based on predictive validities and go back to a 1965 model

sampling study that directly measures the classification efficiency of the then current Army AA composites as compared to FLS composites.

Sorenson (1965) simulated a mobilization population for a Simulation of Personnel Operations (SIMPO) model sampling experiment in which the gain in PCE provided by using FLS composites instead of aptitude areas was evaluated. The means and covariances of the generated scores had expected values equal to those for the Army Classification Battery (ACB) tests in the mobilization population. Predicted performance scores were computed from full regression equations based on the population covariances. Separate validity vectors for eight job families were based on the validities of 55 MOS corrected for restriction in range to provide estimates of job validities in the mobilization population.

The effectiveness of eight two-test composites as assignment variables used by a linear program (LP) algorithm were compared with the effectiveness of full regression equations (FLS composites) using all eleven tests in the ACB. The two-test composites had weights of either 1 or 2. The criterion variables for which validities were available were primarily Army school grades in an era when such grades were normative, reliable, and truly indicative of the soldier's training record. The use of school criterion variables from this era typically provides more dimensionality in the predictor-criterion space and indicates greater PCE than does on-the-job criteria based on ratings. The validities of the two combat aptitude areas (AAs) were, however, computed only against criterion measures based on performance ratings of soldiers stationed in the continental United States.

Twenty entity samples of size 300, thirty samples of size 200, and two hundred samples of size 100 were generated for the model sampling experiment. Appropriate quotas for each job family were used in conjunction with an LP program to assign the entities in each sample to one of eight job families. Assignment was accomplished: (1) first using the AAs as the assignment variables, and (2) a second time using the full regression equations as the assignment variables. The distributions of the MPP standard scores for the two assignment procedures were found not to overlap in MPP scores at all, even for the samples of size 100, and the Army standard score means (mean = 100 and SD = 20 in the unassigned population) were 103 when AAs were used and 107 when full regression equations were used as the assignment variables.

The MPP Army standard score would have equalled 100 if random assignment had been used. Thus, the gain over random assignment is more than doubled by substituting full regression equations for the AAs. Sorenson assumed he had the universe values for

the computation of both the FLS assignment variables and the predicted performance scores used to compute MPP values. His samples were large, but not so large as to completely preclude the presence of correlated error in his assignment and evaluation variables. Repetition of this kind of study using even larger and more recent samples of data for the computation of the parameters of FLS composites is necessary to obtain MPP results that could provide convincing utility findings. As noted earlier, policymakers are not inclined to take action on the basis of MPP gains. The apparent lack of impact the Sorenson findings had in 1965 is attributable, in part, to his not translating MPP gains into utility, and in part to the general pessimism extant at the time; with the all volunteer Army just over the horizon, many believed that the Army could impose little or no control over assignment of new recruits that might conflict in any way with the preferences brought into the recruiter's office by the potential recruit.

C. PSYCHOMETRIC PRINCIPLES FOR PERSONNEL CLASSIFICATION

1. A Taxonomy for Personnel Classification

We define the selection process as the making of decisions in which personnel are rejected or accepted by a potential employer, that is, as the control of entrance into an organization. In contrast, personnel classification is the matching of persons to jobs and placement is the matching of persons to levels within jobs or training programs for those already selected for membership in the organization. When speaking of the Army we will equate "jobs" to military occupation specialty (MOS), and level within jobs to skill level or grade within an MOS. In our usage, classification and placement relate to assignment to an MOS or level within an MOS without regard to designation of a specific military unit or geographical location where "assignment orders" send a specific individual. Assignment is used as a generic term to include decision processes that designate the MOS and grade of an individual, whether the process is classification or placement.

Placement has been given many different definitions in the literature. We are concerned with these definitions primarily because we wish to avoid confusion of placement with classification. Placement is a distinct process, not just a special case of classification when only one measure is being used to assign personnel to jobs. We believe it is important to distinguish between classification and placement, and to be able to use a terminology that permits the consideration of both unidimensional and multidimensional test and criterion sets for all three major processes: selection, classification, and placement.

With regard to placement, in *Psychological Testing* (1988), Anastasia defines placement as assignment to levels within jobs or training programs. She states, "...assignments are based on a single score" (p.189); it is also clear that she would restrict the use of the term placement to the making of personnel assignment decisions with respect to a *single* job, where..."it is evident that...only one criterion is employed, and that placement is determined by the individual's position along a single predictor scale...further that although placement can be done with either one or more predictors, classification requires a multiple predictor whose validity is individually determined against each criterion" (p. 189). We accept her distinction between the focusing on one job or multiple job criteria as the basis for distinguishing between placement and classification. However, we extend both concepts on the predictor side to include both unidimensional and multidimensional processes. Both placement and classification can be based on use of either a single measure or a set of composites to make decisions about matching persons to jobs or job levels.

Cronbach and Gleser's text (1965) utilizes the term placement in a manner entirely consistent with our definition when they refer to personnel utilization procedures that include making assignments to levels of responsibility, to compensation levels within a job, or to difficulty levels in a training program (p. 54). However, they extend their definition of placement to clinical diagnosis and to the selection of alternative treatment of individuals in many situations, including the paroling of prisoners. They include no examples of personnel classification across jobs, defined as above, as a "treatment" in a placement process. Thus, Cronbach and Gleser's concept cannot be used as a precedent for referring to unidimensional classification as a placement procedure.

The desirability of considering differential validity of predictors in selecting test composites to be used for placement to alternative treatments is emphasized by Cronbach and Gleser (1965), "A measure that predicts success under one treatment and not the other would be a much better aid to placement than a measure that predicts both" (p. 59). It is clear that both classification and placement measures are most efficient when test composites possessing differential validity are used instead of a single general measure.

Traditionally, selection has been viewed as a unidimensional process and classification as a multidimensional process. Contrary to this traditional point of view, separate test composites can be used to select for each job family; selection need not be a unidimensional process. Also, classification need not be a multidimensional process. A single test

composite with disparate validities across jobs can be used to accomplish classification across several jobs.

The classification of selected personnel to jobs using a single predictor measure can be effectively accomplished if there is a hierarchical layering across jobs of MPP scores for a job family, and if the single composite is converted to a separate PP score for each job family and these PP estimates are then used as assignment variables. For example, Army AA composite scores can be converted to PP scores based on a single predictor by multiplying the AA composite standard scores (mean of 0 and standard deviation of 1) by the validity coefficient obtained for the given job or job family.

Consider an example in which full least squares (FLS) estimates of job performance are computed as criterion variables. MPP scores for each job family can then be computed as the mean of the FLS estimates. The overall FLS MPP score after assignment could be increased considerably over pure random assignment results by using an assignment process that capitalizes on hierarchical layering. First consider the job family with the highest validities and assign enough of those individuals with the highest PP scores (possibly based on a simple predictor) to this most predictable job to meet the job quota, without regard to their scores on any other assignment variable. Continuing this hierarchical layering process to the job with the next largest validity, then to the next, etc., all persons would be assigned to a job while meeting quotas. This process, a very simple process compared to an LP algorithm, would accomplish optimal personnel assignment if, and only if, there were no differential validity of the composites for their associated job families. If a linear programming algorithm is used to assign the individuals so as to maximize the AA-based PP scores while meeting all quotas, the same result would be obtained--if, and only if, a set of composites with no differential validity were to be used.

Differential validity has been defined by Johnson and Zeidner (1989) as the greater prediction of a criterion by its associated composite as compared with the validity of that composite for performance in other job families--quite a different concept from the hierarchy of validities (or job values) that creates the hierarchical layering effect.

As indicated above, PP scores for each job or job family can be computed by converting a single measure to standard score form and multiplying by the respective validities. In this special case the use of the method for capitalizing on hierarchical layering, as described above, or the use of a conventional LP type optimal assignment, would always yield exactly the same results. The use of a single classification measure

produces a pure form of hierarchical classification, as contrasted to a general classification process which can capitalize on both hierarchical layering and the differential validity among composites.

Just as it is possible to describe an example of personnel classification that is pure hierarchical classification, one can also describe a classification example in which no hierarchical classification effects can exist. We call personnel classification where classification efficiency can occur in the absence of hierarchical layering, allocation. An example of pure allocation is Army's use of the existing AA composites (having equal means and variances in the standardization population) in an optimal assignment algorithm without first creating PP scores or adjusting AA composite scores for the value of jobs to the Army.

Thus, we see that hierarchical classification effects can occur when assignment is across jobs and there are one or more test composites used in an optimal assignment process. However, allocation always requires two or more test composites in the classification battery. Maximum classification efficiency can occur when the assignment algorithm and variables capitalize on both hierarchical layering and differential validities (i.e., when hierarchical classification and allocation are both present in the assignment process).

2. The Role of the FLS Composite in Classification

A set of least squares estimates of the criterion is the most effective set of test composites for use in selection, classification, or placement--if the set uses all the tests in the battery. Such a set of FLS composites cannot be improved with respect to classification efficiency by the elimination of tests that measure only "g", or of any other tests that might reduce the intercorrelations of test composites. These FLS composites are both the most efficient assignment variables and the best criterion measures for evaluating the assignment process.

When each test composite is comprised of a subset of the total classification battery, the same set of composites which maximizes classification efficiency will not usually maximize selection efficiency, or vice versa. When selecting personnel tests for inclusion in the operational battery, or for the further selection of tests for inclusion in test composites, different selection procedures must be followed depending on whether it is *desired to maximize selection or classification.*

3. Brogden's Allocation Model: $(R\,(1\text{-}r)^{1/2})$

Brogden (1951) provides tabled values of the mean predicted criterion scores that would result from the assignment of each individual to his highest criterion score. These criterion variables have zero intercorrelation coefficients among from two to ten job measures. When Brogden's assumptions are met, these tabled values can be multiplied by the average multiple correlation coefficient between predictors and each criterion variable (R), and by a function of the average intercorrelation among the FLS composites designated as predictors (r), to obtain the MPP scores relating to optimal assignment. Denoting the mean criterion score from his table as M, a MPP score is equal to $M\,(R\,(1\text{-}r)^{1/2})$.

Since M is a function of the number of job performance (criterion) variables (that is, the number jobs), the classification efficiency of alternative test batteries for a specified number of jobs is proportional to $R\,(1\text{-}r)^{1/2}$. Thus, a battery of tests could be selected from a set of experimental tests with the objective of improving either R or r as a means of increasing MPP after assignment. If $r = .95$ and $R = .60$, a .05 decrease in r can provide greater benefits than a .05 increase in R (an increment in MPP of .0555 times M contrasted with .0111 times M). It is essential to realize that R and r relate only to FLS composites; Brogden's model cannot be used to estimate the effects of improving R and/or r with respect to any test composite that is not a FLS composite (e.g., Army AA composites).

The MPP score values provided in Brogden's table, and his simple multiplier as a function of R and r, are based on several simplifying assumptions, including the following: (1) All predictor variables are FLS composites; (2) All FLS composites have the same validity (R) and the same intercorrelation coefficient with respect to every other predictor (r); (3) Quotas for each job are equal and the assignments to jobs occur in such large samples that everyone is assigned to his highest FLS composite score; (4) the factor structure of the covariances among the FLS composites corresponds to Spearman's "two factor theory" (i.e., all intercorrelations are explained by a "g" factor although unique factors, one per job, provide additional validity for each job). The robustness of Brogden's tabled results with respect to these assumptions is not known, but his values of MPP as a function of R and r are indicative of an important joint role of these two characteristics of a LSF composite in achieving classification effectiveness.

The least justifiable of these assumptions is the one which affects the increase in MPP with the addition of jobs targeted for assignment. This model assumes a dimensionality of one more than the number of targeted jobs (as does Spearman's two factor model); we know that the dimensionality of the joint predictor-criterion space cannot

exceed the number of predictors in the battery and can usually be expected to be considerably fewer than the dimensionality of the predictor space. Thus, the gain from adding more jobs and composites can be expected to be less in general and to level off much sooner than is indicated by Brogden's model.

4. Potential Classification Efficiency (PCE)

We define the classification efficiency (CE) of a set of test composites in terms of the gain in MPP score under optimal assignment conditions over that obtainable using random assignment. The potential classification efficiency (PCE) of a battery is defined as the gain in the MPP score resulting from optimal assignment obtainable using FLS composites as both assignment and evaluation variables. Brogden's model described above provides a measure of PCE, which, because of his assumptions, is also a measure of potential allocation efficiency (PAE).

The maximum PCE for Army jobs would be obtainable if a separate FLS composite were used to assign to each job. This is, of course, not practical because of the lack of adequate validity data for more than a few of the jobs. Jobs would need to be combined into families in order to provide good estimates of the regression weights for FLS composites even if there were no other reason to create job families. However, PCE is decreased as the number of jobs per family is increased because of the heterogeneity of the jobs in each family is increased. Note that the expected gain from using more FLS composites is expected for reasons entirely different from the rationale provided in the Brogden model; an increased dimensionality is not presumed to occur with the addition of more jobs. Even if there are only three or four independent measures (factors) underlying the covariances of the PP variables for all Army jobs, the use of twenty job families might still be more effective than the current nine job families.

5. The Joint Predictor-Criterion Space

We have noted above that FLS composites can be substituted for criterion scores in accomplishing all computations of validity coefficients and regression weights, and in making any determination of selection or classification effects on performance. Consider the total variances and covariances of the FLS composites as the joint predictor-criterion space. This space is a more limited subset of the total test space and is generally smaller than, but not necessarily entirely included in, the common factor portion of test space in which group factors are traditionally defined. In factor terms, this joint predictor-criterion

space is that part of the factors obtained from total test space by factoring a correlation matrix with units in the diagonals, and extending this solution into the criterion space. The same factor result is obtained from directly factoring the covariances among the FLS composites.

A general factor will usually explain a larger part of the total variance of all factors in the joint predictor-criterion space than is true of total test space, or even of common factor space. When this general factor contribution is spread over most of the independent variables making up the FLS composites, regression equations will be over-determined and there may be many different configurations of regression weights that will provide essentially the same PP scores for a large sample of people. If several sub-samples of examinees are used to compute regression weights, the weight configurations may be surprisingly different while still yielding very similar PP scores for several different sets of weights computed in independent samples. Thus, the presence of overlapping tests in a set of test composites, or a strong "g" component underlying all the measures may make the weights very unstable across samples while still providing as good or better estimates of the criterion as would be provided from a set of variables from which the overlap had been carefully removed.

In summary, a smaller set of relatively independent variables that span the total predictor-criterion space can be readily identified and defined. This derived set of variables will show greater stability in their regression weights across samples. However, these more stable regression weights do not result in these variables providing better estimates of the criterion than is provided by the FLS composites based on the full set of predictors.

6. Optimal Procedures for Classification

The classification efficiency of a set of test composites must be measured in the context of what may be fairly complex operational assignment procedures. Such procedures typically include some provision for considering each individual's differential capabilities for performing well in various jobs while filling job quotas, meeting quality goals for each job family, achieving equal opportunity by race and gender, meeting recruitment commitments, etc. In reality, then, steps planned or taken toward maximizing MPP scores may be doomed not to succeed in such a procedure. Even so, it is only in looking at the CE of a set of composites and the PCE of a battery, under conditions of optimal assignment, that the costs of these various constraints, as well as the costs associated with use of simplified but inefficient composite sets, can be evaluated.

A simple cutting score placed on the score continuum provided by a single selection instrument will maximize selection efficiency. An optimal selection procedure used to assure that all rejected persons have lower predicted performance scores than any persons selected and assigned to a job is feasible but much more complex. A truly optimal selection algorithm requires separate cutting scores on the FLS composites for each job family, rather than a single cutting score on a general measure such as the Armed Forces Qualification Test (AFQT).

Optimal assignment of all selected personnel could be accomplished, without considering constraints, by assigning each recruit to the job family corresponding to his highest test composite score, thus providing the largest MPP score obtainable for a specified set of assignment variables and sample of individuals. The imposition of quotas will of course reduce this MPP score; a reduction in the MPP score is greater when quotas are applied to smaller slices of input (i.e., one week compared to one month).

Traditionally, optimal assignment algorithms, which maximize the mean assignment variable score as the objective function under the constraint of meeting quotas, are referred to as primal solutions. When the primal solution is thus defined, it follows that the dual solution must provide for the minimization of the discrepancy between trial quotas and desired quotas under the constraint of providing a maximum value for the MPP scores (sometimes known as the allocation sum).

When the objective functions of both solutions have been maximized under their respective constraints, they provide the same solution to the same extent that two different primal algorithms could be expected to make the same assignments. A primal solution can be readily made to provide the dual solution parameters (the column constants), and many linear programming "primal" packages offer the dual parameters as an output option.

The mechanisms and consequences of the dual optimal personnel assignment procedure are more readily understood than is true of the primal solution. Consider an array (or matrix) of assignment variable scores in which each row corresponds to a person to be assigned and each column contains the score of an assignment variable associated with a particular job or job family for which the desired quotas are known. When an appropriate value, a column constant, is added to each element of the first column, and repeated until each column array of scores has been adjusted by an additive constant appropriate for each column, each individual can be assigned to the job corresponding to his/her highest adjusted score. Such a set of assignments will meet all quotas and

maximize the mean assignment variables serving as surrogates for predicted job performance. The only difficulty in obtaining such a dual solution is in obtaining the correct set of column constants.

The dual solution provides a basis for intuitively understanding why some jobs have higher MPP scores than others after use of an optimal assignment algorithm. It is clear that the higher the column constant, the more individuals will be assigned to the corresponding job, and the lower their MPP score. If two assignment variables associated with jobs having equal quotas differ with respect to their average correlation with the remaining assignment variables, the one with the lowest such average correlation coefficient will intuitively require a lower column constant and will yield a higher MPP score. Thus, the two operational combat arms job families possessing highly correlated FLS composites and both with high quotas can be expected to have lower MPP scores, even before the possible effects of hierarchical classification are considered.

The advantage gained from the use of an optimal assignment procedure is reduced as the batches are reduced in size. An alternative to the use of small daily, weekly, or biweekly slices of applicants as an assignment batch is to simulate a large sample of synthetic individuals that has the same statistical characteristics as the actual or projected input and, using the known requirements for each MOS, compute the dual solution parameters (i.e., the column constants). These estimated column constants can then be used, one person at a time, to identify the assignment which would maximize the MPP score in the defined population.

Appropriate column constants representing an applicant population can be applied to test composite scores of each applicant to make selection and assignment decisions simultaneously. Rather than selecting on a single measure to provide a pool of recruits who are then assigned to jobs as a distinct second stage, the applicants can be considered for acceptance and use in each job family being considered by the applicant. In such an approach, it can be assured that no one in the rejected group has a higher predicted performance for a job family than anyone selected and assigned to that job family. This approach differs from use of AFQT as a single selection measure. The latter will not preclude the possibility that many in the rejected group will have higher predicted performance scores for a particular job than some newly selected and assigned incumbents.

The algorithm which will accomplish an optimal integrated selection-assignment process is described by Johnson and Zeidner (1989) and called multidimensional screening

(MDS). In this process appropriate column constants are applied to the array of assignment variable scores. The largest adjusted score in each row of the score array is retained; the remaining scores are deleted. The retained scores are then visualized as placed in sort within each column and a cutting score set to accept just enough applicants to meet the job quotas. Actually, the selection-classification decision process can be made, one person at a time, with cutting scores for each applicant; a rank ordering of the PP for each job family within each individual is provided to the counselor or computer charged with making the selection-classification decision.

The use of variable cutting scores derived from the MDS algorithm can provide a more efficient selection procedure only if assignments are made on a non-random basis. However, variable cutting scores can offer leverage in another way. Variable cutting scores reflecting the popularity of various jobs and school courses, or of geographical locations, could take advantage of their corresponding selection ratios to improve the predicted performance of such jobs. The existing minimum scores for each MOS could be retained as the basement cutting scores below which the variable minimum cutting scores could not fall.

D. THE ISSUE OF DIMENSIONALITY IN THE JOINT PREDICTOR-CRITERION SPACE

The dimensionality of the joint predictor-criterion space impacts most dramatically on selection and classification policy when either an increase or decrease in the number of AA composites is proposed. Some measurement professionals have recommended reducing the existing nine AA composites to a set of four composites similar to those utilized by the Air Force. Several measurement experts have emphatically declared that the evidence they rely on the most (comparison of predictive validities and path analyses), leads them to believe that a single test composite, a measure of general cognitive ability, would provide more classification efficiency (albeit of the hierarchical classification type) than the existing Army AA composites. Such a position in favor of a single measure is tantamount to claiming that the joint predictor-criterion space is unidimensional.

We do not believe the joint predictor-criterion space for the existing ASVAB exceeds 3 or 4, depending on the cut-off point for factor contributions selected, even though we do not, at this time, favor reducing the number of AA composites below 9. We seriously believe some increase in classification effectiveness could be obtained from increasing the number of job families and corresponding test composites. In general,

anyone recommending the use of K composites must logically believe that the dimensionality of joint space does not exceed K. However, when K is greater than three, they can logically believe that the dimensionality is less than K but greater than 2.

1. General Factor versus Group Factors

The concept of a dominant general factor was introduced by Spearman. The Spearman "two factor" theory called for explaining all the reliability coefficients and intercorrelation coefficients among tests by recourse to a factor shared across all tests and a factor unique to that test. To provide a matrix of intercorrelations that demonstrates the "two factor" model one must avoid including three or more tests that are essentially parallel forms. The inclusion of only two parallel forms would result in a "couplet" which is traditionally ignored.

The Brogden (1951) model was based on a "two factor" solution of the variance/ covariances among predicted performance estimates with the further assumption that all diagonal elements were equal to one value (R^2) and all intercorrelation coefficients equal to another value (r). A "two factor" model has one general factor explaining the intercorrelation among the PP variables, and one unique factor for each PP variable to explain the remaining reliable variance.

A later generation of factor analysts, exemplified by Thurstone, favored the use of group factors without much attention given either to unique factors or to the general factor. The former were diminished by selecting the tests for inclusion in such a way that at least three somewhat similar tests would be included and by the substitution of communalities for either unity or reliabilities in the diagonals of the correlation matrix. The factor solution would thus occur in the "common" factor space defined by the group factors. A further rotation to a meaningful position of the coordinates (factors) would then be accomplished. The selected set of rotated factors would typically spread "g" over the group factors, thus eliminating "g" as a separate entity.

While the concept of "g" was being de-emphasized in the United States, the influence of the Spearman "two factor" model was still being felt in its entirety in Britain. The Spearman model continued to influence American selection research, but with emphasis being placed more on the unique or specific factor than on "g". The prevailing view over many years was that tests used as predictors of job performance had a large unique component which was situation-specific, and that new empirical data were needed to validate a test for each situation.

Collections of validity data showed wide variations in magnitude for the same or equivalent tests. Despite his desire to find general traits, Ghiselli (1959) finally succumbed to the doctrine of situational specificity. Validity coefficients were believed to be specific to the situation in which they were determined and thus were not applicable to other situations, which would differ in location, time, period, job content, organizational content, background variables, and the interactions of these situational variables. A major challenge to this view is posed by the validity generalization movement.

2. Validity Generalization versus Job Specificity

Schmidt and Hunter (1977, 1981) and Hunter and Schmidt (1982) developed a Bayesian statistical model for testing the hypothesis that variations in validity coefficients in different studies were attributable to statistical artifacts. They found that most of the inconsistent findings across studies were the results of sampling error and failure to take into account other systematic effects such as error of measurement in criteria and predictors and restriction in range. A different view began to emerge--a view of validity generalization--validities could be extended to new situations.

Since the late 1970s, the validity generalization model has been applied to sets of validities in dozens of different occupations, in rejection of the concept that the validity of ability tests was job specific. These results have generally been well accepted by the scientific community. However, the older view of employment testing is so firmly entrenched in scientific thinking that questions continue to arise concerning the methodology of validity generalization and the new results and conclusions that emerge from its application. Schmidt, Perlman, Hunter, and Hirsh (1985) respond to these concerns in a 100 page question-and-answer debate in *Personnel Psychology*. Readers interested in knowing what the continuing technical and philosophical concerns are will find this article quite helpful.

3. A Single General Cognitive Aptitude Measure

It is not difficult to believe in the existence of moderately correlated aptitudes possessed in varying degrees by different individuals while believing in the all-pervasive, overriding dominance of a single general cognitive ability measure in the joint predictor-criterion space. For example, a given task might be performed by one group of individuals using a work style which brings to bear their verbal ability in the use of a handbook, while another group might perform this same task by relying on their ability to perform arithmetic

reasoning and numerical operations, and still another group might accomplish the same task over the same range of outcomes using their ability to visualize mechanical situations and utilize rote memory. Placing all three groups in the same analysis sample would reveal the importance of a general cognitive ability measure as the predictor of task performance.

While the above hypothetical example may not occur frequently enough to explain why more aptitudes can be identified in a test battery than in a set of predicted performance measures, the empirical data do indicate that something that would have a similar effect is, in fact, occurring. There is less dimensionality in the predictor-criterion space than in the predictor space and the sufficiency of a simple measure is more credible in the joint space. To the extent that a general ability measure dominates selection, relevant evidence is, of course, found in the joint predictor-criterion space.

4. Three Theories of Classification Efficiency

A preliminary Navy technical report by Schmidt, Hunter, and Larson (1988) describes another facet of the validity generalization movement whose proponents frequently show a complete dependence on predictive validities and related path analysis approaches to evaluate the effectiveness of test composites in a classification context. An analysis of the Schmidt et al. report is included here because of their discourse on three aptitude and ability theories and their assumption that the comparison of predictive validity results relating to these three theories adequately plumbs the depth of classification efficiency found in the ASVAB.

The authors place their report in the context of:

> an increase in interest "in comparing the relative power of general mental ability and narrower cognitive aptitudes in the prediction of real world performance. This question has important implications for theories of human cognitive abilities. If narrower abilities add nothing to prediction over general ability, then the status of narrower abilities within theories of ability will have to be reconsidered. In addition, (the report) it has important practical implications for personnel selection and *classification*... (p. 1, emphasis is ours).

Schmidt et al. claim that a theory they define in terms of predictive validity (i.e., specific aptitude theory) is the foundation of the differential assignment of personnel to jobs in the military. We will later define a "differential assignment theory" that we believe is the actual provider of this foundation. Their point of view is expressed in context as follows:

> Recent research by Hunter (1983; 1984; 1985) based on very large military samples appears to indicate that general cognitive ability is as good or better

> a predictor of performance in training in most military job families as ability composites derived specifically to predict success in particular job families. These findings are contrary to *the current theory that is the foundation of differential assignment of personnel* to jobs in the military. *That theory*, differential aptitude theory (or *specific aptitude theory), postulates that* specific aptitude factors assessed by particular tests or by clusters of tests *make an incremental contribution to the prediction of performance* over and above the contribution of general cognitive ability. (pp. 1-2, emphasis is ours).

They then discuss their three theories, one of which appears to be *specific aptitude theory* and the other two *general aptitude theory and general cognitive ability theory.*

It appears to us that all three theories, as described and discussed, pertain primarily, if not solely, to predictive validities of test composites based on: (1) job specific measures, (2) general factor specific measures, or (3) a general cognitive ability measure (a "g" factor in the joint predictor-criterion space?). Nowhere is it suggested that one might compare these three theories in terms of the effect on mean predicted performance (MPP) that would be expected to result from the use of an optimal personnel assignment algorithm in conjunction with each theory. Only differences in predictive validity as predicted by each theory are considered by Schmidt et al. in comparing the credibility of these theories (i.e., two of the three theories) in the discussions and results presented in the report.

5. Differential Assignment Theory

We introduce a fourth theory in order to relate the concepts and results presented in our report to the three theories of Schmidt et al. This fourth theory focuses on classification efficiency as measured by mean predicted performance (using a specified assignment procedure). Any loss or gain in predictive validity is relegated by the underlying mathematics (a result, not an assumption) to what may in many cases be a minor role in the achieving of an increase in classification efficiency. In this discussion we will call this fourth theory "differential assignment theory." The concepts and postulates of this theory are described in some detail earlier in this chapter and further elaborated in another report by Johnson and Zeidner (1989).

The report provides data comparing results obtainable from the application of specific aptitude theory to those obtainable from general cognitive ability theory, but does not discuss differential assignment theory. In the Navy report the "QVT" validities are not factor validities but are instead validities for a four-test subset of the ASVAB (each factor defined as a composite of ASVAB tests). Since these validities do not pertain to factor

validities and thus relate to the "general aptitude theory," only two of the three theories discussed in the report are actually compared empirically.

If two sets of test composites, neither of which is an FLS composite, are compared in order to see which set will provide the largest mean predicted performance (i.e., classification efficiency) under optimal assignment conditions, the set of composites showing the smaller average predicted validity could easily be the one with the greater classification efficiency. This hypothetical situation is what one would expect to occur if one set of composites were created to maximize predicted validity and the other to maximize classification efficiency.

Schmidt et al. do not appear to be supporters of the specific aptitude theory. We probably oppose this theory, as defined by the authors, more strongly than they do. The substitution of the goal of achieving an increment in the predicted validity of job specific composites for the achieving of an increase in MPP in the context of optimal assignment can seriously interfere with the selection of classification efficient tests for inclusion in a battery and the forming of classification efficient composites whenever the composites are not FLS measures. The present ASVAB and present Army AA composites have largely resulted from the pursuit of the erroneous specific aptitude theory. Even the placement of jobs into job families has most probably been adversely affected by an undue consideration of the specific aptitude theory.

We agree with Schmidt et al. that the Army aptitude areas, as currently used, are of questionable value. However, we do not believe that the ASVAB is irredeemable. We believe that considerable classification efficiency is potentially obtainable from the existing ASVAB if it is used in accordance with differential assignment theory. The ASVAB would be even more promising if its development had not been largely based on the erroneous specific aptitude theory--a reliance on efforts to increase predictive validity rather than on application of differential assignment theory. As a result, the current ASVAB has much less classification efficiency than would a battery developed specifically to maximize classification efficiency.

We are convinced that the contribution of Hunter and Schmidt and several others in pointing out that the obtaining of classification efficiency is not an easy task is very important. We also believe that the achieving of classification efficiency is a more difficult task than many assume and cannot be achieved without a specific effort aimed at developing classification efficient predictors, batteries, and selection-classification

procedures. Our goal is to convert this perception of the difficulties to be overcome into constructive effort rather than into a hopeless pessimism that would kill all possibilities of future progress.

In summary, the theories and results provided in the Schmidt et al. report, although claimed to be highly relevant to the value of the ASVAB for classification, are based entirely on predictive validity and thus have little relationship to classification efficiency. From the discussion in this report it appears that Schmidt et al. believe that their results provide overwhelming evidence on the utility of using FLS composites as compared to the utility of using a single measure of general cognitive ability to accomplish classification of personnel. We believe that they are using a grossly inadequate measure of classification efficiency that can shed very little light on the contribution that the use of FLS composites could provide in a classification context.

6. Compatibility of Validity Generalization and Classification

We believe Schmidt et al. described in their "specific aptitude theory" a modus operandi that is followed by many researchers in the military services and industry. While not an appropriate approach for use with respect to classification batteries, it is one which is often used, i.e., Schmidt et al. were not tilting at windmills. Those who have read earlier reports by Schmidt and Hunter, with a variety of junior authors, in which differential assignment theory has been correctly identified and used (Hunter and Schmidt, 1982; Schmidt, Hunter and Dunn, 1987), have to be surprised that they did not also address it in Schmidt et al. (1988), although we have noted their general preference for citing predictive validity and path analysis results when criticizing the ASVAB.

We see no reason why the presence of PCE in a battery takes away from either the validity or the importance of the validity generalization concept or of the application of the Bayesian statistical model which has contributed so much to our understanding of the validity of general cognitive ability measures across jobs. We also see no reason why our acceptance of the all-pervasive presence of a general cognitive ability measure in predictive validity data should make us pessimistic about the future usefulness of classification batteries.

E. EVALUATING ALTERNATIVE PERSONNEL UTILIZATION POLICIES

Most would agree that the ultimate objective of military personnel utilization, a process that includes selection, placement, and classification, is to maximize the effectiveness of each military unit in accomplishing its mission. Ideally, the measure of utility associated with alternative personnel utilization policies would reflect this ultimate objective. Unfortunately, a direct measure of unit effectiveness is prohibitively expensive. Developing such an ultimate criterion would require determining the effect of many alternative personnel configurations on the unit's accomplishment of important missions; research would have to be accomplished separately for each kind of unit. Thus, it is essential to identify substitute criteria that are sufficiently relevant, or valid and affordable, if objective evaluations of management tools are not to be precluded.

Our approach to utility assumes linear relationships between a number of variables that form a chain linking predicted performance to a dollar criterion. Included in this chain is the presumed linear relationship between the performance measures and the productivity of individuals on each job. We believe predicted performance can be appropriately used as a surrogate of productivity. Predicted performance, in turn, is assumed to have a linear relationship with the value of that performance to the organization. It is also assumed that value can be appropriately expressed in dollars. It is preferable to convert productivity to dollars separately for each job, but it is usually necessary, and fortunately justifiable, to settle for the conversion of mean predicted performance across all jobs to dollars.

In this section we examine assumptions made for the processes proposed for evaluating alternative assignment policies. In general, any means of determining "payoff," expected productivity, system benefits, or utility of each person-job match outcome, using an optimal assignment algorithm, can also be used in the aggregate, as a means of comparing alternative personnel utilization policies. However, it is highly desirable to compare alternative policies using a metric that can be directly compared with cost. The utility of each policy can then be expressed as the difference between benefits and costs. We propose that the value of being able to trade off benefits and costs as a means of evaluating the utility of proposed policy changes justifies the use of feasible (i.e., not too expensive to use) benefit measures that may be less valid theoretically but provide this capability.

F. CHALLENGES TO THE USE OF VARIOUS COMMON METRICS

There are those who may challenge the feasibility of comparing, across jobs, objective measures of benefits attributable to the assignment of a person to one job instead of another, as is essential in a classification process. Similarly, some may deny the legitimacy of expressing utility as a function of benefits and costs. Such critics claim that the value of performance in military jobs resulting from an assignment process cannot be expressed in the same metric used to measure the costs of recruiting, training, and distributing personnel of the prescribed quality to military jobs.

Some critics may oppose the use of criterion variables constrained to have a normal distribution or a linear relationship to predicted performance. Others may object to the conversion of the productivity or value metric into dollars. Among the reasons given for not using a normally distributed dollar metric based on a linear relationship to predicted performance include the following:

(a) The military services are not profit-making organizations.

(b) The capability of achieving military objectives cannot be priced in dollars, e.g., what are lives worth?

(c) The avoidance of catastrophic failures is much more important than most achievable increases in mean predicted performance.

(d) Many crews or units will not have their effectiveness increased by adding more high quality personnel beyond some small percentage of the total strength.

One concern expressed from time to time through the years in the military context is that analyses of the peacetime force may provide results that differ from an analysis of the force in war, i.e., the effective garrison soldier may not be the effective combat soldier. By necessity, most analyses are not analyses of combat. But there is no compelling argument for the proposition that proficiency and effort are not the best predictors of later performance, even in combat. Additionally, the value of tasks to be included in a performance measure is generally judged in the context of combat scenarios.

Thus, productivity gains may not win the war though they may contribute to its outcome. The goal of military selection and classification utility research is to increase productivity of the military work force through providing better individual or team performance at lower costs.

G. A TAXONOMY OF CRITERION MEASURES

Approaches for obtaining numerical values to represent utility can be divided into two major types: those that separate estimates of costs and benefits; and those that integrate costs and operational constraints into a single utility variable. This division is shown at the top stem of our taxonomy tree in Figure 1.1.

Further branches of our utility taxonomy within each of these two divisions are shown in Figures 1.2 and 1.3. The first branching in each figure is based on whether the basic sources of data are performance measures or expert judgments. For the separate estimate of costs and benefits, as shown in Figure 1.2, the branch of utility measures that is based on predicted performance contains the most commonly used techniques. Conversely, for the approaches shown in Figure 1.3--all of which directly provide utility measures reflecting a merger of benefits, costs, and operational constraints--expert judges are the source of the data for the more commonly used methods.

The clearest example of using objective measures to obtain an integrated utility value for a policy, without separately estimating benefits and costs, may be provided by a field experiment in which an existing operational policy is compared with alternative policies for which similar data have been collected. This is the left main branch of the utility categories shown in Figure 1.3. An example can be provided by a field experiment in which utility at the unit level is defined as an integrated function of: (1) readiness or other measure of effectiveness, and (2) depletion of resources as a result of unit efforts to achieve effectiveness. The overall, integrated, measure of utility is "unit readiness." A separate measure of costs is not necessary since depletion of resources, a cost type variable, is reflected in the unit's readiness to deploy and engage the enemy.

H. AIR FORCE PAYOFF APPROACHES FOR DETERMINING UTILITY

Over the last two decades the Air Force (AF) has conducted a large number of studies to develop and refine methodology for estimating the "payoff" values associated with alternative operational policies. Even more studies have been conducted to model AF personnel acquisitions, training, promotion, and assignment policies. The AF operational system currently uses a policy specifying technique.[1] (Ward, 1977; Ward, Pina, Fast, and Roberts, 1979.)

---

[1] The person-job match module (Promis-PJM) of the AF assignment system.

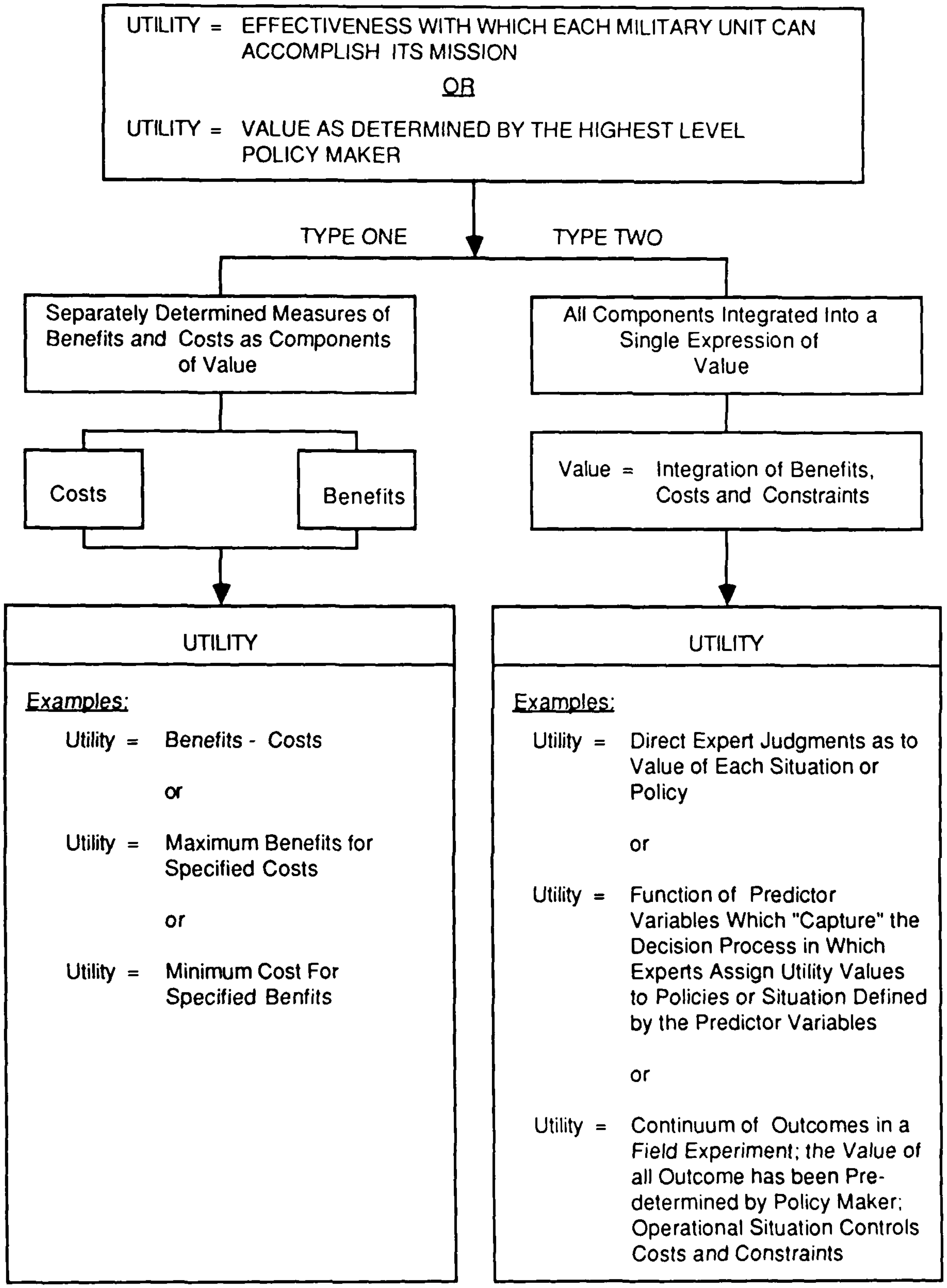
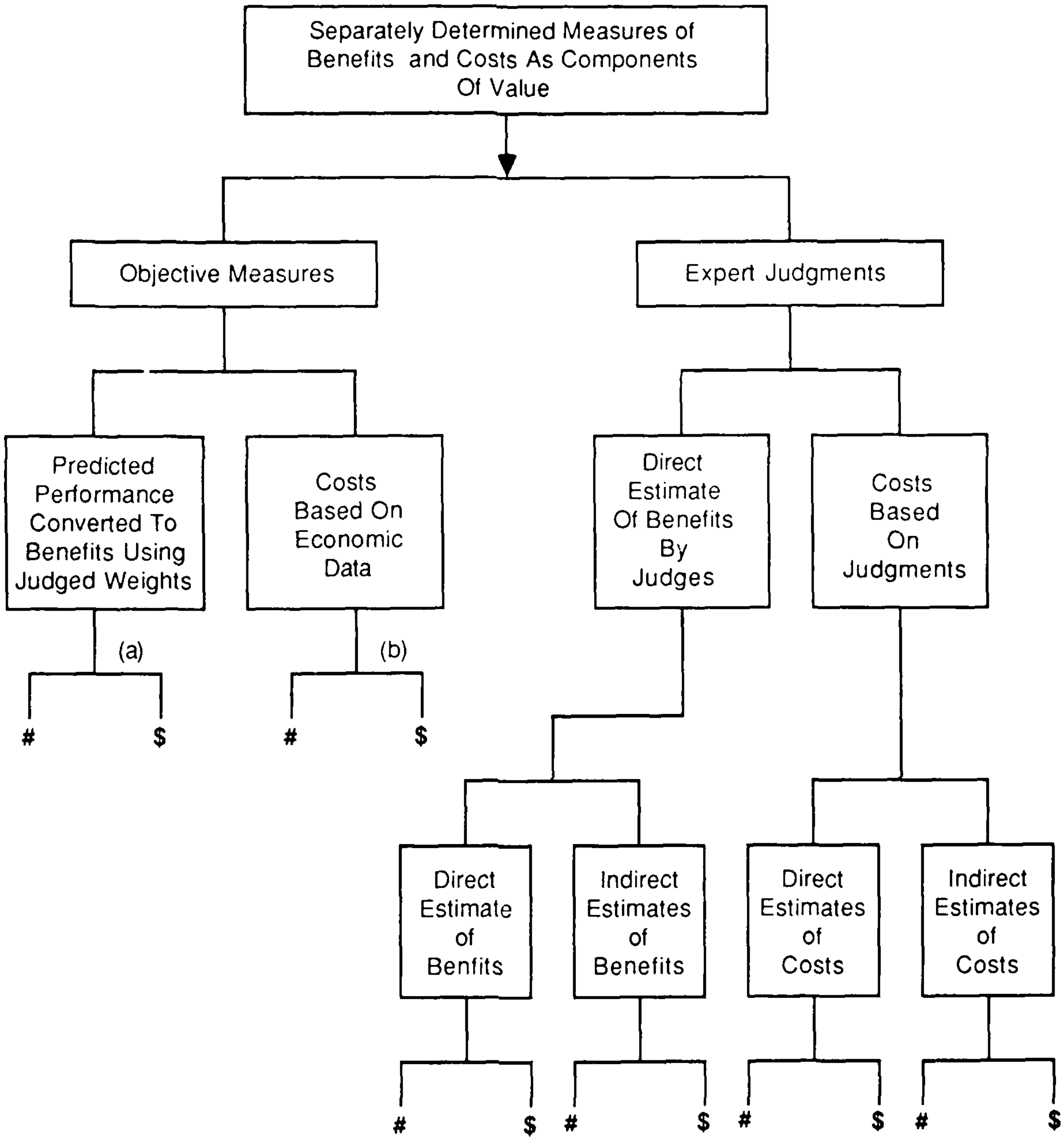


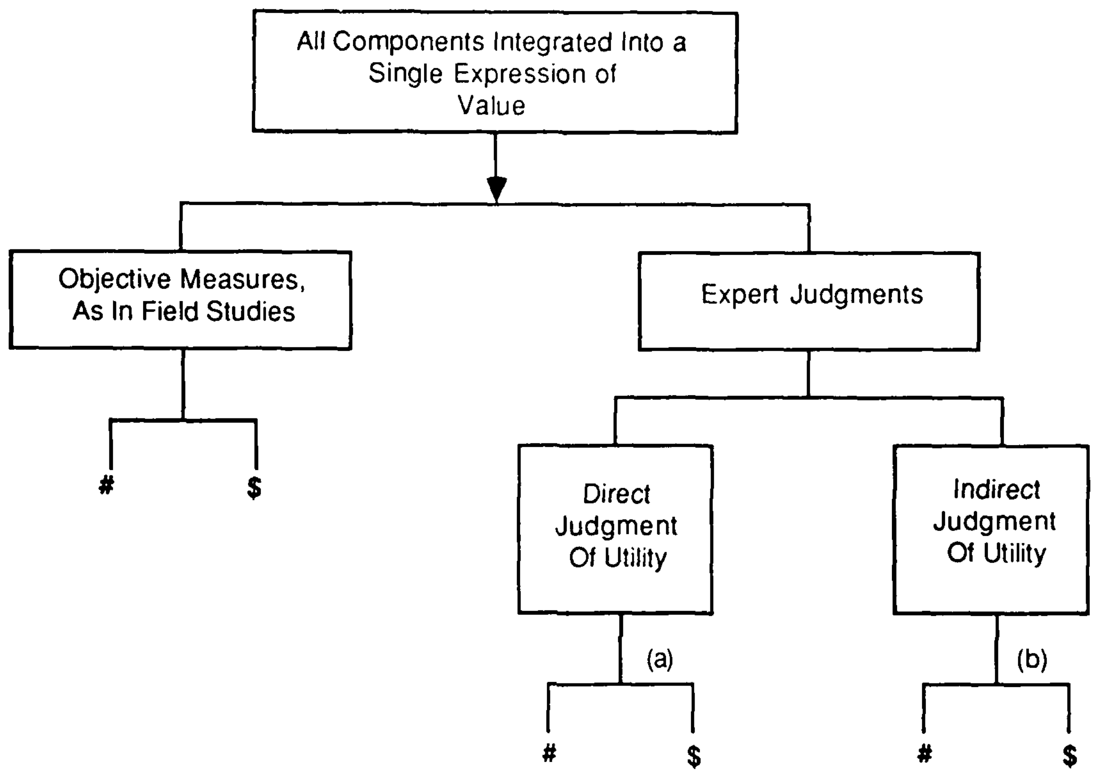
Figure 1.1 A Major Branching in a Taxonomy of Utility Measures



a $ = Dollar Criteria

b # = Other Value Metrics

Figure 1.2. Division of Separately Ot ned Benefits and Costs Into Objective and Subjective Measures ir llars[a] or Other Value Metric[b]



a $ = Dollar Criteria

b # = Other Value Metrics

Figure 1.3. Division of an Integrated Value Measure Obtained Through Field Studies or Expert Judgment in Dollars[a] or Other Value Metric[b]

The AF research teams have investigated a variety of techniques for identifying and modeling policies that explain the attempts of judges to place value on the results of operational decisions.[2] They have conducted studies on the use of explicit value weights that ascribe utility to personnel system outcomes. Weights applied to variables descriptive of persons, jobs and/or operational situations are made by expert judges to reflect their contribution to utility or "payoff". Consideration was given to a similar technique in which judges of "payoff" directly assign utility points to predictor variable intervals--the point allocation method.

The research teams gave more serious attention to the implicit determination by judges of the utility of personnel utilization situations. They investigated policy capturing, hierarchical policy specifying and combinations of the two. Either technique could be used to accomplish a task of great interest to us, the estimating of the value of jobs in conjunction with converting predicted performance to predicted benefits.

In one policy-capturing approach, an expert judge assigns a value to sets of person-job matches that are defined in numerical values of variables descriptive of the person, the job, and the operational situation. Least squares regression equations are then computed to fit the data in which the judged criterion values corresponding to each situation are the dependent values, and the descriptive variables are the independent variables.

One difficulty Ward found with the use of policy capturing was that one set of experts was needed to properly consider "management-related" variables (such as in situational progress in filling classrooms), and another set required to judge "quality-of-assignments-related" variables (such as those related to matching persons to jobs where they will perform well and/or be satisfied) (Ward, 1977). Ward stated that "there might be judges who can adequately combine the management-related information and there might be judges who can handle the quality-of-assignments-related-information. But it was felt that it would be difficult to identify policymakers who could appropriately combine both types of variables into an acceptable policy through the policy-capturing process" (p. 7).

---

2 The references in Ward's 1977 review include ten separate authors and co-authors of AF reports explaining, exploring, evaluating, and applying payoff functions. There has been a sizable number of additional reports on the application of the hierarchical policy specifying model to AF problems. Ward's preface cites a dozen others as having made major contributions to "policy specifying models". Ward has been the central figure in the conduct of this impressive institutional effort. Christal provides an explanation and demonstration of policy capturing in a journal article (1968).

Hierarchical policy specifying requires the expert judge to consider pairs of the descriptive predictors (the same predictors as described in connection with policy capturing) in conjunction with the payoff value of all pairs of predictor values. This process provides a description of the surface defined by each pair of predictors and the utility value. A mathematical algorithm is required to aggregate these pairwise responses into a decision hierarchy. While it would be possible to obtain a mathematical estimate of the utility surface in hyperspace over all variables from an integration of the pairwise estimates--if, and only if, all pairs can be judged by at least one expert--this is not the way it is done in the AF operational implementation. The final form of the specified policy used in the AF operational system (PROMIS/PJM) is a combination of functions for each of the pairwise utility surfaces formed into a policy hierarchy. A utility measure in this form is adequate for making operational decisions, but would be awkward for use in evaluating alternative policies.

I. CONVERSION OF PERFORMANCE TO BENEFITS

The expert judgment process in the AF policy specification model permits the expression of non-linear relationships between the predictors and payoff or utility, and permits relationships among pairs of variables that require cross-product terms and higher order terms (e.g., quadratic, cubic, quartic, etc.) to plot the resulting surfaces. It would be possible to adapt the AF model for use with job performance, productivity, and mission variables as descriptive variables in order to produce a value for the contribution of each job--in lieu of the payoff values for particular decisions or situations--as the model output.

These job values could then become the multipliers of the mean predicted performance scores for each job before aggregation into the mean predicted benefit over all jobs. For most management studies, it is often appropriate to assume that such a benefit measure is linearly related to a dollar value. It would then be appropriate to use the SD_y approach to convert the weighted predicted performance values, the benefits, to a dollar criterion that permits trading off costs and benefits.

Alternatively, if predictive performance and the value of the contribution of each job proved to be significantly non-linear, this non-linear relationship between performance and job contribution could be retained and reflected in a non-normal distribution of benefit scores. These benefits scores would then be aggregated across jobs to provide the mean benefit score. SD_y would then be used as above or modified to reflect a further non-linearity to create the dollar criterion.

J. OBJECTIVE VERSUS SUBJECTIVE MEASURES

The designation of an *ultimate* criterion implies that the expert making this designation is the ultimate authority against which there is no appeal. There can be a different ultimate criterion for each set of such authorities. The scientists must accept the policymaker's decision as to the nature of the benefit or utility measure he wishes to be optimized, since he is the only relevant (i.e., the ultimate) authority on what constitutes the most credible criterion. Thus, we will not render an opinion as to whether estimates of utility based on an objective measure of productivity and costs, or the opinions of the top policymakers will provide the truest or most credible measure of utility. We will confess to feeling more comfortable with utility measures that are maximally based on empirically derived measures of both performance and costs.

We see many advantages to the separate determination of benefits and costs using a comparable metric as a means of determining the best of alternative policies for recruiting, selection, classification, training, and retention. We suspect most policymakers would have more faith in the use of empirically obtained relationships with objective measures of performance and with objective measures of cost than they would have in the use of judgments made by experts on the utility of person-job matches that impact on both predicted performance and operational considerations such as meeting class room quotas. Thus, were we to believe that the relationship of performance to the value of performance on jobs were seriously non-linear, we would prefer to convert predicted performance to a benefits score separately for each job or job family, permitting the benefits variable to assume any appropriate mathematical curve. We would find such an approach more credible than reliance on expert judgments to determine "payoff."

We see no way of avoiding expert judgments if it is desired to make a determination of the value of the contributions of job incumbents at different levels of predicted performance. The descriptive variables for use in either policy capturing or policy specifying models should include difficulty of tasks, importance to accomplishment of unit mission of good performance, criticality of failures, availability of supervision, and many other variables. The first step of a policy-specifying study used to make a determination of values as described above would be to identify the appropriate variables.

K. THE ASSUMPTION OF LINEARITY

The basic selection and classification utility equation depends only on linearity. Indeed, behavioral science research in general almost always utilizes this assumption.

Hunter and Schmidt (1982) provided a detailed analysis of the question of meeting this statistical assumption. Their analysis of this assumption should be read by those concerned with the questionability of linearity. Brogden and Lubin's work described by Hunter and Schmidt (1982), in attempting to identify non-linear predictor-criterion relationships in large military samples is also worthy of note. They conclude that not one of the non-linear equations cross-validated successfully in a new sample; the non-linear functions were never superior to simple linear functions. They state:

> Thus, it appears that an obsessive concern with statistical assumptions is not justified. This is especially true in light of the fact that for most purposes, there is no need for utility estimates to be accurate down to the last dollar. Approximations are usually adequate for the kinds of decisions that these estimates are used to make (VanNaerssen, 1963, p. 282; Cronbach and Gleser, 1965, p. 139). Alternatives to use of the utility equations will typically be procedures that produce large errors, or even worse, no utility analysis at all. Faced with these alternatives, errors in the 5%-10% range appear negligible. Furthermore, if overestimation of utility is considered to be more serious than underestimation, one can always employ conservative estimate of equation parameters (e.g., r_{xy}, SD_y) to virtually guarantee against overestimation of utilities. (pp. 245-246)

When referring to the predictor-criterion correlation in utility models it is necessary to use a proxy criterion, performance, in place of a dollar-valued criterion since the latter is not available for direct measurement. The assumption is then made that both criteria are linearly related to the predictor and to one another. Schmidt, Hunter, McKenzie and Muldrow (1979) believe the relationship between the proxy and the dollar-valued criteria to be linear, or if not, the results underestimate utility because ceiling effects will lower the correlation between the predictor and the proxy criterion. It is unlikely that any production function will continue to increase linearly with ever-increasing increments of high aptitude employees: the law of diminishing returns eventually applies. For example, if there are too many high aptitude employees, they will be assigned to tasks of less value to the organization and thus affect utility estimates. However, in real world situations, with regard to most performance-aptitude distributions, the linear relationship between aptitude and value appears to be a reasonable assumption.

L. OUR APPROACH

The desirability of converting MPP into job benefit measures using a separate conversion for each job hinges on the acceptability to policymakers of the assignment of disparate importance values to jobs as recommended by Hunter and Schmidt in numerous articles. If management is willing to provide disparate values for different performance

levels of incumbents within a job, they should also be willing to permit the consideration of different values across jobs for use in an operational system that is making assignments. An assignment process making use of these disparate job values would greatly increase the utility of the personnel utilization process. Thus, any study of the utility obtainable from personnel utilization processes that does not consider the possibility of disparate job values is almost certainly providing an underestimate of utility.

In the simulation described in Chapter 3, MPP was not converted into benefit measures reflecting importance or value separately for each job, because the use of such information in an operational system would require a major policy change that may never actually be made. Also, the large research effort required to obtain importance data for making separate SD_y estimates specific to each job would entail resources well beyond the scope of the simulation project reported in Chapter 2.[3]

If policymakers are willing to assign importance values separately across jobs and/or within a job in the future, MPP can readily be converted to reflect disparate job or job level values. Alternatively, rational estimates of SD_y, such as the global estimating procedure, could be used to obtain values separately for each job. As indicated in a later chapter, the type of SD_y estimating procedure employed would not effect the selection of the procedure producing the greatest benefit. There is a large, ongoing Project A effort to obtain ability level/performance values within and across jobs. We recommend the desirability of each service evaluating its productivity gains while making their own assumptions and estimates. If this were to be done, Project A job importance values may readily be employed and utility results compared with the results of models of utility that make simplifying, more affordable, assumptions.

We see considerable advantage in the separate measurement and consideration of benefits and costs, using a common metric such as dollars, as a means of evaluating alternative policies relating to recruiting, selection, classification, training, and retention. We find more credibility in estimates of utility based on objective measures of performance, and would thus prefer to use expert judgments only in the provision of values for the contributions of job incumbents. We further believe the linearity assumptions made by Brogden in his classical utility function (1949), and in the elaboration of this function by

---

3 Those readers who are unfamiliar with the use of the SD_y procedure to convert performance into productivity gains in dollars can find a detailed discussion of this approach in an earlier report (Zeidner and Johnson, 1989).

Cronbach and Gleser (1965), are appropriate for use in most investigations of classification utility. Finally, we support the investigation of utility using a variety of analytical, simulation, and field studies; we certainly urge the use of simplifying assumptions when the alternative would be no utility study at all. We also believe these simplifying assumptions will frequently be justified by the extremely small amount of bias they inject and the savings in both time and money the use of these assumptions provide to the investigator. Chapter 3 provides a detailed description of a model of the acquisition and allocation of human capital within the context of classical economic production theory used in our simulation.

M. STUDY DESIGNS: DETERMINING CLASSIFICATION EFFICIENCY

1. Mean Predicted Performance

Classical functions for selection utility are based on first obtaining the mean predicted performance (MPP) score for the selected group. Similarly, a value of MPP can be obtained for those optimally assigned to jobs, and when this assignment is also accomplished using FLS composites, MPP is a measure of potential classification efficiency (PCE) for the operational battery. The effects of selection and classification, whether accomplished sequentially or simultaneously, can be measured in equivalent terms through the use of MPP. Thus, the use of MPP as a measure of process efficiency is the thread that links selection and classification utility.

If one is willing to assume that all joint distributions are normal, it is easy to describe a selection and/or classification process in terms of multiple definite integrals of a multivariate normal curve. The numerical solution of this function is MPP, and if possible to obtain, would be a very useful result. Unfortunately, it is difficult, if not virtually impossible to obtain a numerical solution of such a multiple integral when there are more than three jobs. For this reason, the determination of PCE involving classification of personnel to more than one job requires, for all practical purposes, the simulation of the optimal personnel assignment process and the computation of MPP scores for the personnel assigned to each job.

2. Simulation Using An Empirical Sample

A simulation experiment to determine the MPP scores corresponding to alternative classification policies requires access to realistic predictor scores and knowledge of the

relationship of these predictor variables to performance. Predictor scores may be obtained from data banks which contain either all of the members of a specified population, or a random sample of that population. Ideally that population would be the youth population, or would at least contain all applicants. However, the population of those entering military service is adequate for determining the PCE or classification efficiency (CE) of operational instruments and policies, provided it is not desired to evaluate a policy that includes assigning an input group based on lower selection standards than were in effect when the data were collected.

A simulation experiment to determine either PCE or CE in terms of MPP requires a definition of both the assignment composites--the variables used in the optimal assignment algorithm as the objective function to be maximized--and evaluation composites--the predicted performance (PP) variables that are based on all available information. These PP variables are the FLS composites referred to earlier in this chapter and are used to compute MPP scores of the individuals assigned to jobs in the experiment. Both sets of variables, the assignment and the evaluation variables, should be computed using weights that have been computed on data that is independent of the individuals in the data bank supplying predictor scores for the simulation. If the weights used to compute assignment and evaluation variables are not assumed to be the actual universe values, but are instead computed on samples from the prescribed universe, the two sets of variables should be computed on independent samples.

3. Simulation Using A Sample of Synthetic Scores

Predictor scores can also be generated by a model sampling technique that produces synthetic scores with the statistical properties of samples drawn from a population having a known covariance matrix and scores having multivariate-normal joint distributions. The expected covariance matrix for a sample of synthetic scores is equal to this universe covariance matrix. Such model sampling concept and implementing procedures are described in considerable detail by Johnson and Zeidner (1989).

Model sampling simulations have the advantage of being able to provide random samples drawn from any population for which a covariance matrix is either available or can be estimated using such statistical techniques as restriction in range corrections. Thus, a reasonable estimate of a youth population can be provided, extending the range of personnel utilization policies that can be simulated. Also, any number of random samples,

of any desired size, can be readily generated, making it possible to provide independent samples for cross validity or other research designs.

Simulations based on scores drawn from empirically derived data banks have the advantage of more precisely reflecting the actual score distributions. For example, Army input is not only curtailed at the lower end as the result of selection, but is also systematically censored in the upper part of the score distributions of predictor variables. While it is easy to mirror the curtailment, the censoring provides a more difficult, although not impossible, challenge to the investigator who is using model sampling techniques.

N. THE ROAD AHEAD: PAVING THE WAY FOR CHAPTERS 2-6

The scope of this chapter was restricted to those operational problems and associated psychometric principles, prior results, and theoretical issues that relate to the efficient use for classification purposes of the existing ASVAB. The next chapter, true to their delineation of our topic, will describe the classification system, EPAS, that is expected to become fully operational in FY1989-90. The third chapter describes simulation and utility analysis that provide answers to several methodological questions relating to the fine tuning of EPAS. This simulation and associated utility analyses provide one of the very few, if not the only examples (at the moment), of this important research approach.

Chapter 4 continues the theoretical discussion of psychometric principles begun in Chapter 1. The results of the simulation described in Chapter 4 are interpreted in the context of these principles and prior results. Immediate operational implications of research conclusions drawn from the simulation are identified and recommendations made for either immediate implementation or impact studies by management analysts.

In Chapter 5, the most promising operational changes in the classification systems of the military services are identified on the basis of psychometric principles and prior results. New research, some already initiated but incomplete, is described as necessary to confirm the utility gains we believe these changes would provide. Chapter 6 provides a road map for operational implementation; the schedule for addressing the proposed changes by service researchers and management analysts is provided.

CHAPTER 2. THE ARMY MANPOWER PROCUREMENT AND ALLOCATION SYSTEM

Edward J. Schmitz and Roy D. Nord

There have been many recent applications of testing to enhance the productivity of organizations. Examples of such applications, found throughout an earlier report (Zeidner and Johnson, 1989, July) generally provide an illustration of the substantial benefits that could result to an organization through selecting its applicant population with cognitive tests. However, very few of these analyses have resulted in actual organizational implementation of alternative testing procedures.

The military establishment presents the most complete example of a personnel system that uses testing to make key personnel decisions. This chapter describes how the U.S. Army, the largest single organization in the country, operates its manpower management system, particularly with respect to selection, classification, and allocation. The sections of this chapter describe: the structure of the Army's personnel system, including the tests used for selection and job classification; current personnel policies and procedures with respect to initial personnel placement; current operational systems for executing these policies; enhancements to the current operational system that will improve organizational productivity attributable to testing procedures; and future enhancements that can be made in the operation of the personnel system.

A. THE ARMY'S PERSONNEL MANAGEMENT SYSTEM

Table 2.1 outlines the personnel flow of the enlisted force. The first step in the personnel system, recruiting, is managed by the U.S. Army Recruiting Command (USAREC). USAREC operates a field "sales force" of approximately 5,000 recruiters across the country. These recruiters are responsible for finding sufficient numbers of qualified individuals to enlist in the Army each year.

The first step in formally applying to the military is to take the Armed Services Vocational Aptitude Battery (ASVAB). The ASVAB takes approximately three and

one-half hours to administer. This battery is designed to assess the individual's eligibility for the military and trainability for various occupations. The ASVAB is given to applicants at 70 Military Entrance Processing Sites (MEPS), and associated Mobile Examining Team Sites (METS), and to many high school juniors and seniors.

Table 2.1. The Enlisted Personnel Flow

Personnel State	Number (average number 1973-81)
Applicant	330,000
Qualified Applicant	222,000
Contract	168,000
Enlistment	162,000
Completed Enlistment (Nonattritee)	100,560
Reenlistment	24,000
Completed Second Enlistment	22,500

The ASVAB, used both to select and classify applicants through various combinations of tests, is comprised of ten separate tests. Table 2.2 identifies the ten tests, the abbreviation for each, and the reliability of each as reported by McLaughlin, Rossmeissl, Wise, Brandt, and Wang (1984).

One key part of the ASVAB is that which comprises the Armed Forces Qualification Test (AFQT). Four of the tests (AR, MK, PC, and WK) are combined to determine the AFQT, used to determine enlistment eligibility, along with eligibility for various enlistment incentives. The AFQT is scored in percentile terms, normed against the 1980 youth population. Individuals scoring below 10 are legally prohibited from military service. An individual presently needs a percentile score of 16 or above to be eligible to join the Army (Army Regulation 601-210). The percentile ranges are typically collapsed into test categories for administrative purposes. Table 2.3 identifies these test categories and their corresponding percentile ranges.

Table 2.2. Tests Comprising ASVAB Versions 8/9/10

Subtest	Abbreviation	Reliability
General Science	GS	0.86
Arithmetic Reasoning	AR	0.91
Paragraph Comprehension	PC	0.81
Word Knowledge	WK	0.92
Numerical Operations	NO	0.78
Coding Speed	CS	0.85
Auto Shop Information	AS	0.87
Mathematical Knowledge	MK	0.87
Mechanical Comprehension	MC	0.85
Electronics Information	EI	0.82

Table 2.3. Percentile Score Ranges of the AFQT Test Categories

AFQT Category	Percentile Score Range
I	93 - 99
II	65 - 92
IIIA	50 - 64
IIIB	31 - 49
IV	10 - 30
V	0 - 9

Those individuals found eligible on the basis of AFQT are then evaluated as to whether they are qualified for enlistment on the basis of other criteria. In addition to the ASVAB, they are given a physical examination and screened for other prerequisites such as

education, criminal background, and attributes that may be required for a particular job (e.g., citizenship, driver's license, typing ability).

Once qualified for military service, the individual meets with a guidance counselor to sign the enlistment contract. The military occupational specialty (MOS) available to recruits--258 in all--in which training will be given is selected at this time, along with other enlistment options. The standard enlistment is three years of active duty. High school diploma graduates who score in test categories I-IIIA (AFQT 50 or above) are eligible for an enlistment bonus of up to $8,000 or special educational benefits if they agree to enter a difficult-to-fill MOS. The enlistment bonus may entail an active service commitment of four to six years; the amount of the educational benefits also depends on the length of the enlistment: the longer the term, the greater the level of benefits.

As mentioned earlier, classification and assignment to MOS is performed prior to the applicant's acceptance into the military and usually several months prior to entering active duty. The Army's current classification system also uses the ASVAB. The ASVAB is comprised of ten separate subtests, four or five of which are combined into nine aptitude area composites to determine job eligibility. Each of the 258 entry-level MOS uses one or more of the aptitude area scores to determine eligibility. The qualifying score required to permit training in a particular MOS ranges from 85 to 120. The aptitude area composites are also normed against the 1980 youth reference population with a mean of 100 and a standard deviation of 20. Each entry level job requires an aptitude area score above a predetermined level. Table 2.4 identifies the tests used to calculate AFQT and aptitude area scores.

Few individuals enter the Army directly after signing the enlistment contract. Most are placed into the Delayed Entry Program (DEP), before they are called up to active service; many wait for a period of up to 12 months, so that they may complete school or wait for the starting date of their specific training course. Approximately five percent of DEP contracts eventually fail to meet their enlistment commitment. While many of these individuals could be prosecuted for failing to fulfill their contracts, few are ever charged by the government. The expense and adverse publicity, and the expected poor performance of such individuals, if forced to serve, make it undesirable to do so in peacetime. The outcome of the entire procedure is that only about one half of all formal applicants actually are enlisted in the Army.

Table 2.4. Tests Comprising the Army's Aptitude Area Composites

Aptitude Area Composite	Abbre-viation	AR	AS	CS	EI	GS	MC	MK	NO	PC	WK
Clerical	CL	X						X		X	X
Combat	CO	X	X	X			X				
Electronics	EL	X			X	X		X			
Field Artillery	FA	X		X			X	X			
General Maintenance	GM		X		X	X		X			
Mechanical Maintenance	MM		X		X		X		X		
Operators/Food	OF		X				X		X	X	X
Surveillance/Communications	SC	X	X				X			X	X
Skilled Technical	ST					X	X	X		X	X
AFQT		X							X	X	X

Recruits receive both basic military and occupational training on entering the Army. Basic training lasts 8 weeks, while initial job training may be from 6 weeks to a year depending on the MOS.

New recruits exhibit substantial turnover. Nearly 20 percent fail to complete their initial year of service, and over 30 percent do not finish the typical 3-year enlistment. Many of those that do complete their initial enlistment tour are ineligible for reenlistment because of inadequate performance as measured by job skill tests, rank attained, and supervisor assessments. Of the soldiers eligible to reenlist, only about one third do so.

Soldiers who reenlist usually make a career of the Army. Such soldiers have passed through a double screening process: a preference for continued service and adequate performance as judged by the Army. Inadequate performers will have been screened out by the Army's personnel policies during the first term. Both the military and the individual have strong economic incentives to remain beyond the reenlistment point. The retirement system offers immediate annuities of 50 percent of basic pay after 20 years

of service, providing a strong inducement for the individual to remain until the twenty year vesting point.

From the Army's standpoint, considerable resources in acquiring, training and developing career soldiers have been invested. Also, for most Army jobs, there are no corresponding civilian jobs where the Army can find the skilled labor it needs. Thus, few soldiers are screened out for poor performance after the first reenlistment is completed.

B. PERSONNEL SYSTEM OBJECTIVES

The Army operates a complex human resource planning system, such as that described by Niehaus (1979). Figure 2.1 illustrates five basic components of such a system. First, inventories of requirements and personnel available to fill requirements are maintained. These requirements and inventories are evaluated against one another to determine how to distribute the projected personnel and forecast the additional supply needed to fill remaining requirements. The organization then executes the plan through its personnel system, and periodically evaluates whether the system is operating in equilibrium. If the system is not achieving its objectives, alternative policies are considered to bring supply and demand into balance.

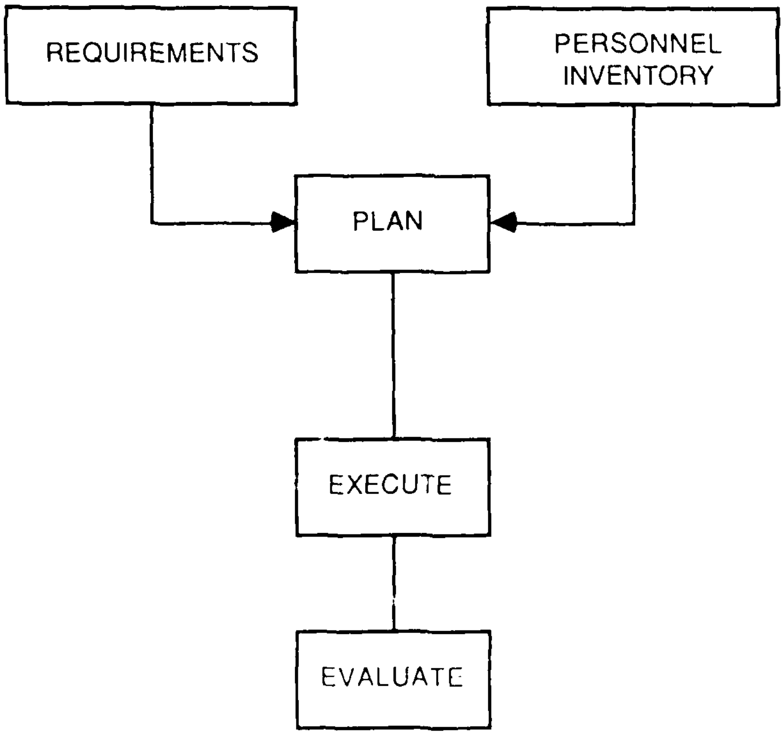


Figure 2.1. The Human Resources Planning Model

The nature of the Army's personnel system places great emphasis on the efficient management of its personnel system. The Army must train and develop all its personnel, as very few enlisted jobs exist for which the Army can find people with transferable training. (The medical area is one of the few exceptions.) Hence, the Army must rely on internal supply to fill virtually all of its requirements for experienced manpower. The only recourse open to the Army is to recruit, train, and retain people without prior experience.

The model that the Army relies on to maintain its enlisted personnel in balance is the Military Occupational Specialty Level System (MOSLS). MOSLS assures that the training requirements, personnel strength, budget, promotions, and recruiting objectives are in agreement for all jobs in the Army. The design of MOSLS is described by Eiger, Jacobs, Chung, and Selsor (1988). There are both budgetary constraints on the cost of its enlisted force and personnel constraints upon the total number of people in the Army. Also, the training base further restricts personnel placement.

MOSLS provides the precise estimates of requirements needed to operate a detailed personnel allocation system. If a organization such as the Army could not determine the detailed vacancies that need to be filled, there would be substantial mismatching of training resources with manpower. MOSLS determines the requirement for new recruits, a requirement produced by MOS and training class. To minimize training costs, the Army generally schedules classes evenly over the course of the year.

The Army spends over 600 million dollars in recruiting efforts. An additional 1.5 billion dollars is spent on initial skill training of enlistees. Moreover, enlistees receive over 5 billion dollars in pay and allowances during their first term of service. In total, over seven billion dollars are spent to maintain the Army's first term enlisted force at the current level of productivity. The major organizational decisions that determine what jobs recruits will have for the next three years have been made before they leave the MEPS. Or, to put it another way, to change the decision of the type of training means no cost at the recruiting point; however, once the recruit has entered training many thousands of dollars will have been committed.

Thus, the focus of testing is particularly relevant in screening the 130,000 new recruits accepted by the Army each year. The service can exercise the most flexibility and opportunity in applying information about predicted performance on the applicant group.

Two major personnel objectives drive the Army's personnel acquisition system: filling requirements and maintaining a high level of productivity. Meeting total numerical

personnel requirements is the dominant objective of the Army's job allocation system. The Army is saddled with the largest and most complex recruiting problem of all the military services. Over 40 percent of 325,000 military enlistees enter the Army each year. Also, compared to other military services, the Army provides relatively few marketable skills, and those that are marketable are often performed under undesirable working conditions.

A survey of job satisfaction among 18-21 year old youths found the military ranked significantly below the labor market in general, and the Army was by far the lowest ranked service (Blair and Phillips, 1983). The Gates Commission in its study of ending the draft predicted the Army would face the greatest difficulty because of nonpecuniary factors associated with its working conditions (Studies for the Commission on the All Volunteer Armed Forces, 1970).

Yet the Army obtains its required number of new soldiers each year despite the magnitude and difficulty of the Army's recruiting task, achieving or nearly achieving its numerical recruiting objective every year of the all-volunteer force except one. (See Table 2.5). No significant shortfall in meeting the recruiting objective has been reported since Fiscal Year 1979.

However, the Army desires the most productive recruits possible to enter the service in terms of low attrition and high quality of job performance. Two principal indicators of soldier performance have emerged with respect to post-Vietnam manpower management objectives.

Attrition generally is defined as the failure to complete the initial enlistment. High attrition is often an indicator of poor individual performance or of poor management performance. Reducing attrition can significantly raise Army productivity in a number of ways. First, recruits who attrite fail to produce any substantial output. Nevertheless, the Army has invested considerable recruiting and training resources to develop soldier skills. On the average, over $16,000 is invested in recruiting and training soldiers without prior service who attrite.

A second cost of attrition is its impact on supervisors. First-line supervisors spend a major portion of their time providing on-the-job training to new recruits once they enter units. Poor performers require substantially more supervision and discipline, leading to lower performance for the unit.

Table 2.5. Recruiting Trends

Fiscal Year	Non-prior Service Objective	Non-prior Service Accessions	Percent of Objective	Percent Diploma Graduate	Percent Test Score Category I-IIIA	IV
1974	184,700	182,224	98.7	50.1	52.5	17.8
1975	183,900	184,600	100.4	57.8	57.6	10.0
1976	180,200	180,175	100.1	58.6	54.8	7.6
1977	167,900	168,398	100.3	59.2	34.2	43.8
1978	126,900	124,029	97.7	73.7	37.9	39.3
1979	149,200	129,284	86.7	64.1	30.6	46.0
1980	157,800	158,179	100.2	54.3	26.0	51.9
1981	116,800	117,915	101.0	80.3	40.0	30.9
1982	115,600	120,353	104.1	86.0	53.0	19.2
1983	132,400	131,702	100.3	87.6	61.4	12.0
1984	131,353	131,702	100.3	90.8	63.4	10.2
1985	119,000	119,121	100.1	90.7	62.9	8.5
1986	126,875	127,143	100.2	90.8	63.0	3.8
1987	119,500	120,512	100.8	91.1	66.7	3.9

Finally, attrition detracts from the overall efficiency of the Army by increasing the size of the training pipeline. To maintain the same number of productive soldiers in units, an Army with a high level of attrition requires more soldiers in training. Each attritee occupies a nonproductive training space--a training slot that will not lead to a productive soldier. Even if training incurred no cost, an Army with high attrition would need to maintain more soldiers in the training base than an Army with low attrition.

To achieve low attrition the Army has sought to recruit high school graduates, who tend to be much more likely to complete their enlistments successfully. Research by Buddin (1981, 1984), Baldwin and Daula (1984), and Manganaris and Schmitz (1984)

found that high school graduates have about one-half the attrition of nongraduates during the first tour of duty. Table 2.5 shows the increase in the proportion of high school graduate recruits in the post-Vietnam era. Since FY80 the percentage of recruits with high school diplomas has increased from 54 percent to over 90 percent.

The relationship between AFQT and attrition has been observed in many studies in various services over the years (e.g., Navy personnel, Sands, 1978; Air Force, Flyer, 1956; and Marine Corps, Goodstadt and Glickman, 1975). In the Army, some more recent evidence exists that first-term attrition is related to AFQT. Eaton and Nogami (1980) and Manganaris and Schmitz (1985) found individuals in higher AFQT categories had lower attrition than those in lower mental categories. AFQT percentile scores have been found to be negatively and significantly related to lower attrition (Buddin, 1981; Baldwin & Daula, 1984). However, in contrast, Belgrave and Nogami (1986) found black males in test categories IIIB and IV had the lowest rates of attrition among all groups.

The second measure of productivity is job proficiency. The ASVAB predicts a criterion of job proficiency best. High ASVAB scores are desired because they are highly related to training success and job performance (Zeidner, 1987). Armor, Fernandez, Bers, and Schwarzbach (1982), McLaughlin et al. (1984), and Fernandez and Garfinkle (1985) have found that AFQT and aptitude area scores predict job performance as measured by Skill Qualifications Tests. Nelson, Schmitz, and Promisel (1984) and Scribner, Smith, Baldwin, and Phillips (1986) found AFQT to predict critical task performance in antiaircraft defense and tank gunnery. The Congressional Budget Office (1986) reviewed soldier performance and acknowledged that there is a substantial body of evidence indicating a general relationship between military productivity and mental ability scores, although there are many unanswered questions about the nature of that relationship, such as the relationship of individual performance to group capabilities.

Table 2.5 provides two measures of ASVAB scores for Army recruits commonly used as indicators of quality: the percent of recruits scoring 50 or above on the AFQT (test categories I-IIIA) and the percent in the lowest eligible test category (IV). It should be noted in examining the distribution that there were norming problems with the ASVAB for all services during FY77-80 (Maier & Truss, 1983). Nevertheless, since FY81 the proportion of recruits in test categories I-IIIA has increased from 40 percent to two-thirds, while the share of recruits in test category IV has declined from 30 percent to under four percent.

In addition to setting objectives for the AFQT, education, and gender composition of the new recruits, the Army sets distributional goals in terms of the AFQT for each job. These quality goals assure that the population of recruits entering each MOS has at least a certain number scoring above average (test categories I-IIIA), and no more than a specific percent of those in the lowest eligible test category (IV).

Improvements in soldier performance have required the commitment of substantial additional resources. In order to be able to select and assign individuals with higher expected performance the Army has had to expend substantial additional resources. Individuals expected to perform well tend to be sought by industry and universities. To improve the Army's ability to acquire candidates likely to be good performers has required pay raises, increased recruiting resources, larger enlistment bonuses, the creation of special scholarship programs (the Army College Fund), and shorter enlistment tours. It has been estimated by U.S. Army Recruiting Command, Daula and Smith (1986), and Dertouzos (1985) that it costs from four to eight times as much to acquire a male high school graduate in test category I-IIIA as a graduate in category IIIB or IV.

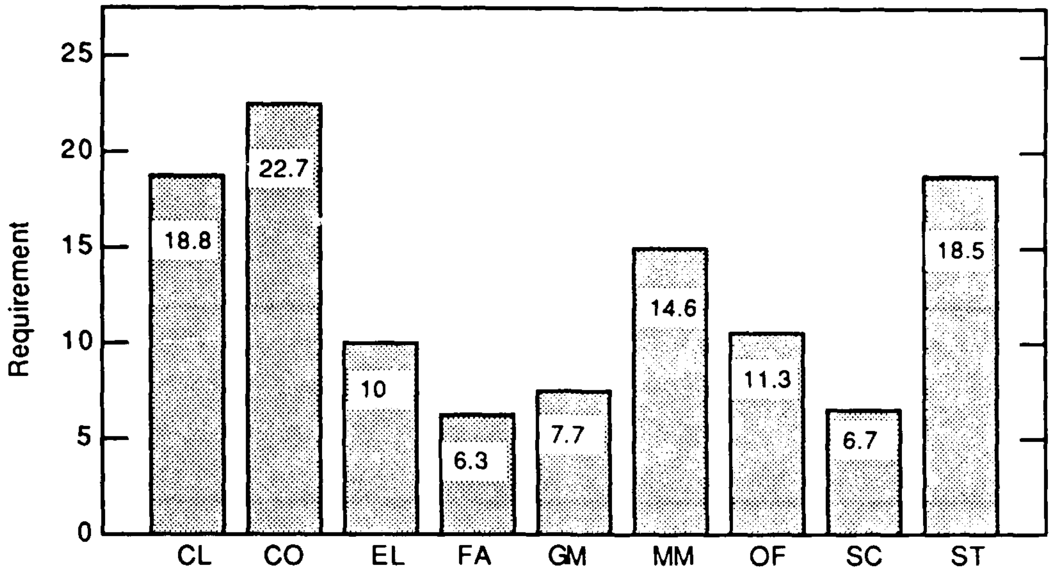
C. OPERATIONAL SELECTION AND ASSIGNMENT SYSTEMS

A complex system of explicit standards, sophisticated computer systems, management goals, and judgments produce the Army's selection and assignment systems. How these systems operate with respect to the organization's policies and the use of testing information is described here.

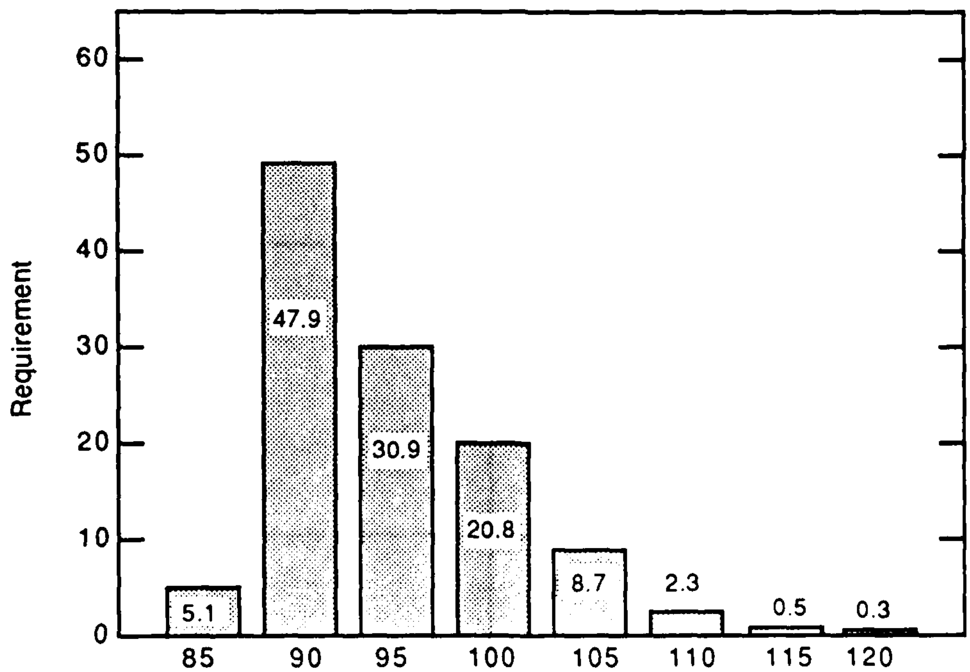
The allocation of recruits to MOS is controlled by three separate operations: MOS enlistment standards, "switch settings", and guidance counselor presentations.

The Army imposes minimum qualifying scores on the aptitude area composite associated with each MOS. Figure 2.2 shows the distribution of qualifying scores and their distribution across aptitude areas. The average aptitude area qualifying standard score is about 94, or somewhat below the mean standard score of 100 for the youth population.

The second action to control the distribution of recruits is the determination of which MOS will be open to a particular type of recruit at any given time. Allocation of personnel is manually controlled by analysts at USAREC who must frequently modify the list of which jobs are open. These "switches" are operated at the MOS level on the basis of educational level, Armed Forces Qualification Test (AFQT) score category, and gender.



(a) FY 86 Requirements by Aptitude Area



(b) FY 86 Requirements by AA Cut Score

Figure 2.2. 1986 Army Manpower Requirements by Aptitude Area and AA Cut Score.

An MOS is determined to be either open or closed for an individual based upon these three factors. This mechanism is used to assure that the AFQT category composition of each MOS is within the goals established by the Army for the year. For example, a recruit may not be offered an MOS in a particular area even if he or she is well qualified for it if he or she falls within an AFQT category that is in excess for that MOS.

The final part of the process is a computer program (Hierarchy) that recommends specific MOS to new recruits. All three military services operate similar programs (Kroeker and Rafacz, 1983; Ward, Haney, Hendrix, and Pina, 1978). This program, part of the Recruit Quota System (REQUEST), first eliminates those MOS in which the applicant is unqualified, then examines the current fill requirements of the remaining MOS. The final step in the process is the creation of an ordered list of up to 25 MOS which reflects the payoff to the Army of each potential applicant-job match. In the present system this ordering is exclusively determined by the priority and fill requirements of the MOS that are open to the applicant.

The USAREC guidance counselor, together with the applicant, uses this list to determine the MOS in which the individual will train. Although the applicant can request to see other MOS for which he or she is qualified, the guidance counselors are generally successful in negotiating individuals into MOS that represent critical needs for the Army. In FY87 we found that over 75 percent of all contracts signed were for a person-MOS match that was among the Army's 25 top priority or critical jobs.

The resulting system is successful in meeting several important goals. It does an acceptable job of filling requirements. For example, in FY84 the system met total accession requirements, 90 percent of the individual MOS training requirements, and 88 percent of the MOS quality goals.

However, the present system does little to improve job performance. The reason for this is that the emphasis upon filling training seats tends to "crowd out" consideration of factors affecting job performance or attrition: aptitude area score, education, gender, and AFQT. Table 2.6 lists the factors in the Hierarchy program and their weights. All the weights are on factors associated with filling training seats; no weights are given to factors that predict job performance. This is the result of two policies: the high priority on filling requirements and the operational use of AFQT as the indicator of job performance.

Table 2.6. MOS Ordering Factors in Hierarchy

Factor Type	Factor	Weight
MOS	Cohort Seats	32%
MOS	Requirements	26%
MOS	Training Seats	17%
MOS	Class Seats	15%
MOS	Priority	10%
Applicant	Aptitude Area Score	0
Applicant	AFQT	0
Applicant	Gender	0

In practice AFQT is used as both the selection and classification instrument. The use of the AFQT distribution goals helps the Army avoid many undesirable outcomes in the assignment process. However, AFQT provides very little job matching capability--classifying recruits into those jobs they will perform best. The aptitude area scores were constructed as the most valid predictors of performance for MOS (McLaughlin, Rossmeissl, Wise, Brandt, and Wang, 1984). To the degree that there are significant differences among jobs in terms of where an individual could be expected to perform best, the aptitude area composites can generate differential performance. That is, if any given individual could be assigned to an MOS in an aptitude area in which he could be expected to perform best on the basis of his test scores, then performance could be increased over a system in which assignment is largely random.

The reason for the relatively low assignment efficiency is the reliance on low job standards for a high quality recruit population, combined with the focus on filling requirements. The average recruit today qualifies for 85 percent of the jobs in the Army. The average test category I-IIIA recruit qualifies for 96 percent of all jobs. Thus, the aptitude area scores have little influence on directing the placement of recruits into MOS in which they are most highly qualified.

In addition, research by Manganaris and Schmitz (1984) has identified significant differences in the way AFQT category, education, and gender affect attrition rates in different MOS. No attrition information is used in making assignments.

D. DEVELOPMENTS TO IMPROVE THE ASSIGNMENT SYSTEM

Simulations of the optimal assignment of Army recruits to jobs, such as those done by Schmitz and Nelson (1984a, 1984b) and Fernandez and Garfinkle (1985), have indicated that job performance can be increased considerably. However, these batch assignments would be infeasible under present assignment procedures. Applicants must be placed in jobs at the time they negotiate their enlistment contract. This situation in analogous to a class of operations research problems referred to as the secretary problem (Tamaki, 1984). A decision maker must select a secretary from a finite number of applicants. There is no recall; each applicant must be evaluated and either accepted or rejected in sequence. There is no analytical solution for such a problem if more than three jobs and applicants are involved.

Project B, the Enlisted Personnel Allocation System, was developed by ARI to improve the assignment of new recruits to their training MOS while maintaining the same overall enlistment procedures. EPAS uses a four step strategy to assign recruits to MOS. First, forecasts are made of the numbers and types of applicants available for assignment over the ensuing 12 months. Then a plan is developed for the allocation of these recruits over that period. This plan is used to guide the training seat recommendations made to each prospective soldier who is offered an enlistment contract. Finally, the system is frequently updated to assure that the overall plan is in close agreement with current supply and demand.

The forecasting, planning, and classification functions of EPAS are performed by four different modules. Figure 2.3 illustrates how the different modules of EPAS fit together. Forecasts of recruit supply and training requirements are produced from two separate modules. These modules generate inputs to an optimization routine. The optimization first assures that all requirements, targets, and policies are met. Then among those feasible alternatives the optimization finds the distribution of people to jobs that will provide for both the maximum performance and minimum attrition from the pool of available recruits. Finally, the optimization module passes information on the optimal distribution of manpower to the classification module. Applicants are classified on the

basis of how they compare to the best available candidates who are likely to arrive in the current recruiting environment.

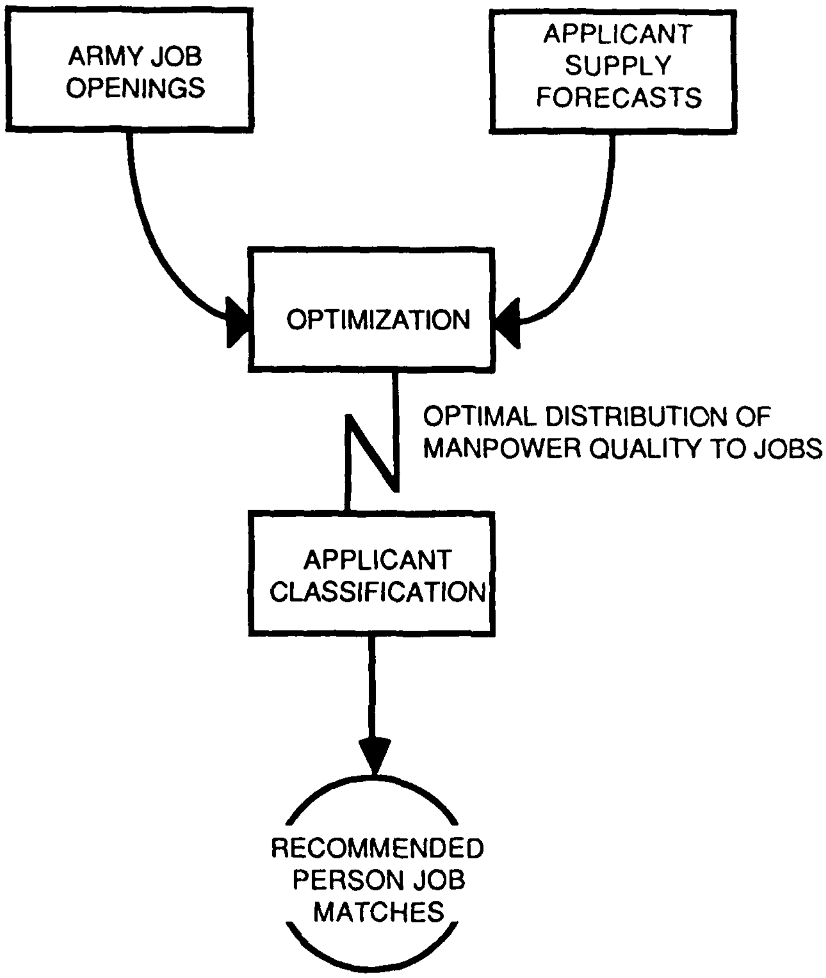


Figure 2.3. EPAS Applicant Classification

Projections of job vacancies made by MOSLS provide the monthly training seats that must be filled over the next year. Aggregate supply forecasts can come from a combination of a supply model such as Horne (1985) for groups that are supply constrained, and USAREC missions for groups that are less difficult to recruit. These forecasts are disaggregated by education, gender, AFQT, and aptitude area score combinations into 81 separate groups. For example, high school graduate males with above average AFQT scores are disaggregated into 31 separate groups according to job

specific test scores. These supply groups provide a categorization that provides a connection between the categories managed by USAREC and a differential performance-based classification system.

Table 2.7 presents an example of two such supply groups. Both groups are identical with respect to recruiting characteristics (graduate males, test category I-IIIA) and would receive identical recommendations under the current allocation system. EPAS would be likely to recommend assignments for the first group in CL, EL, GM, and ST aptitude clusters, while the second group would likely be assigned to jobs in the CO, MM, OF, or SC ap'itude areas.

Table 2.7. Examples of EPAS Supply Groups

Characteristic	Group 1	Group 2	Difference
Gender	Male	Male	–
Education	Graduate	Graduate	–
AFQT Category	I-IIIA	I-IIIA	–
Aptitude Area Scores			
CL	123	111	13
CO	114	126	−12
EL	127	111	−16
FA	119	117	2
GM	124	113	11
MM	115	123	−8
OF	114	123	−9
SC	118	123	−5
ST	124	112	12

Once information is available about the demand for jobs and supply of recruits, a plan is developed to allocate supply to demand. This plan is complicated by the fact that it

must be concerned with not simply filling jobs or achieving performance goals, but must attempt to achieve both goals while satisfying a large number of distribution and timing constraints. For example, neither too many nor too few recruits may be brought into the Army each month, and the distribution of AFQT scores in each occupation over the course of the year must achieve the goals for each occupation. The time dimension is a critical complicating factor that must be included in the problem. Very few recruits enter the Army in the same month they enlist; nearly all enlistments enter the Delayed Entry Program for periods ranging from one to twelve months.

The planning system solves a network optimization problem to determine an optimal allocation plan for each month's expected applicant population. The general form of the problem is:

$$\text{Minimize } Z = (\Sigma)\ c_{ij}N_{ij} \qquad (11.1a)$$

Subject to:

$$(\Sigma)\ N_{ij} <= N_i \qquad (11.1b)$$

$$(\Sigma)\ MOS_{ij} <= MOS_j \qquad (11.1c)$$

$$(\Sigma)\ MOS_i{}^*{}_j <= MOS_j{}^* \ . \qquad (11.1d)$$

The c_{ij} are the costs associated with assigning an individual of type i to job j. N_{ij} is the number of individuals of type i assigned to job j. The first constraint (equation 11.1b) is that the number of productive individuals in each of the 81 supply groups is limited. Equation 11.1c refers to the fact that each MOS must be filled. The final constraint is that the high quality recruits (i*) must be equal to or greater than some specified quota (j*) for each MOS.

The objective is arbitrarily defined to minimize costs, where costs can be defined as some function of the assignment of recruits to jobs. Cost can be defined as related to attrition, job performance, or a combination of both factors. Two measures of performance have been used thus far: technical job performance as predicted by the aptitude area score, and attrition as predicted by characteristics of the recruit and job (Manganaris and Schmitz, 1985).

The network optimization first performs an evaluation of the alternative allocation plans to identify a plan that satisfies the distributional constraints and achieves the least cost (or highest aggregate performance level). The above problem is further complicated by the time dimension--the Army must not only match up applicants with jobs, but it must

schedule their arrival into training at the proper time. The combination of all of these constraints results in a very large problem--12 months × 81 Supply Groups × 258 MOS. Restrictions on time in DEP and class start dates reduces the size of the problem somewhat, but there remain over 50,000 assignment combinations to be evaluated each month.

Since alternatives other than the least cost must also be considered, alternative nonoptimal solutions are also evaluated. Alternatives are evaluated in terms of how much greater the costs increase relative to the optimal level. The closer the alternative is to the optimal, the more highly it is evaluated.

The results of the planning model are then used to guide the job recommendations for each applicant. The planning model generates a list of preferred MOS assignments for each supply group. Alternative assignments are evaluated by how close the alternative is to the optimal assignment.

Figure 2.4 provides an illustrative example of how EPAS would produce optimal guidance. In this example three supply groups are assigned to six MOS. The optimal guidance scores are the relative payoffs for each MOS/time period/supply group match. The "best" or optimal match for each supply group is assigned an arbitrary value of 1000. Other alternatives are scaled in proportion to their "reduced costs" that are estimated from a sensitivity analysis of the optimal solution (Hillier and Lieberman, 1980, p. 195). This provides a way to evaluate the relative desirability of all feasible alternatives, not simply the optimal alternative. For example, in Figure 2.5 the optimal recommendation for a recruit belonging to supply group 10 would be MOS 71L in October or 13B in November. However, if a recruit from this group wished to enter in December, he should be directed towards either 19E (score 910) or 11X (score 900).

Figure 2.5 shows how the results of the planning model are then used to guide the job recommendations for each applicant; this is illustrated on the upper left side of the figure. Each recruit then has this information combined with his specific test scores and other characteristics, along with the MOS status at the time of contracting (lower right corner) according to a payoff function. This part of the decision algorithm is analogous to both the present Army system and procedures used by the Air Force and Navy. The guidance provided by the planning system assures that the many goals and constraints on the distributional aspects of the assignment are met while performance is improved.

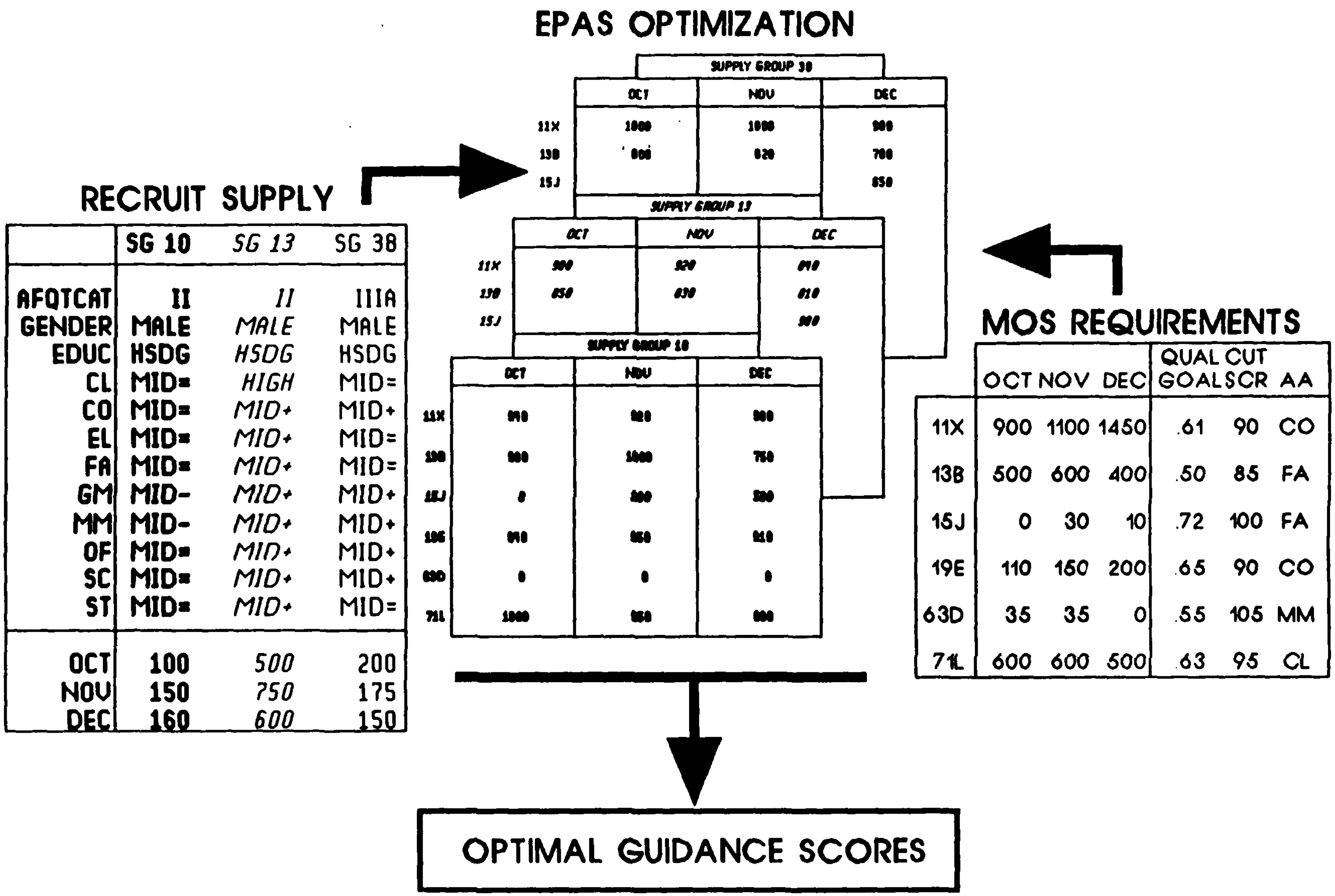


Figure 2.4. EPAS Optimal Guidance Flowchart

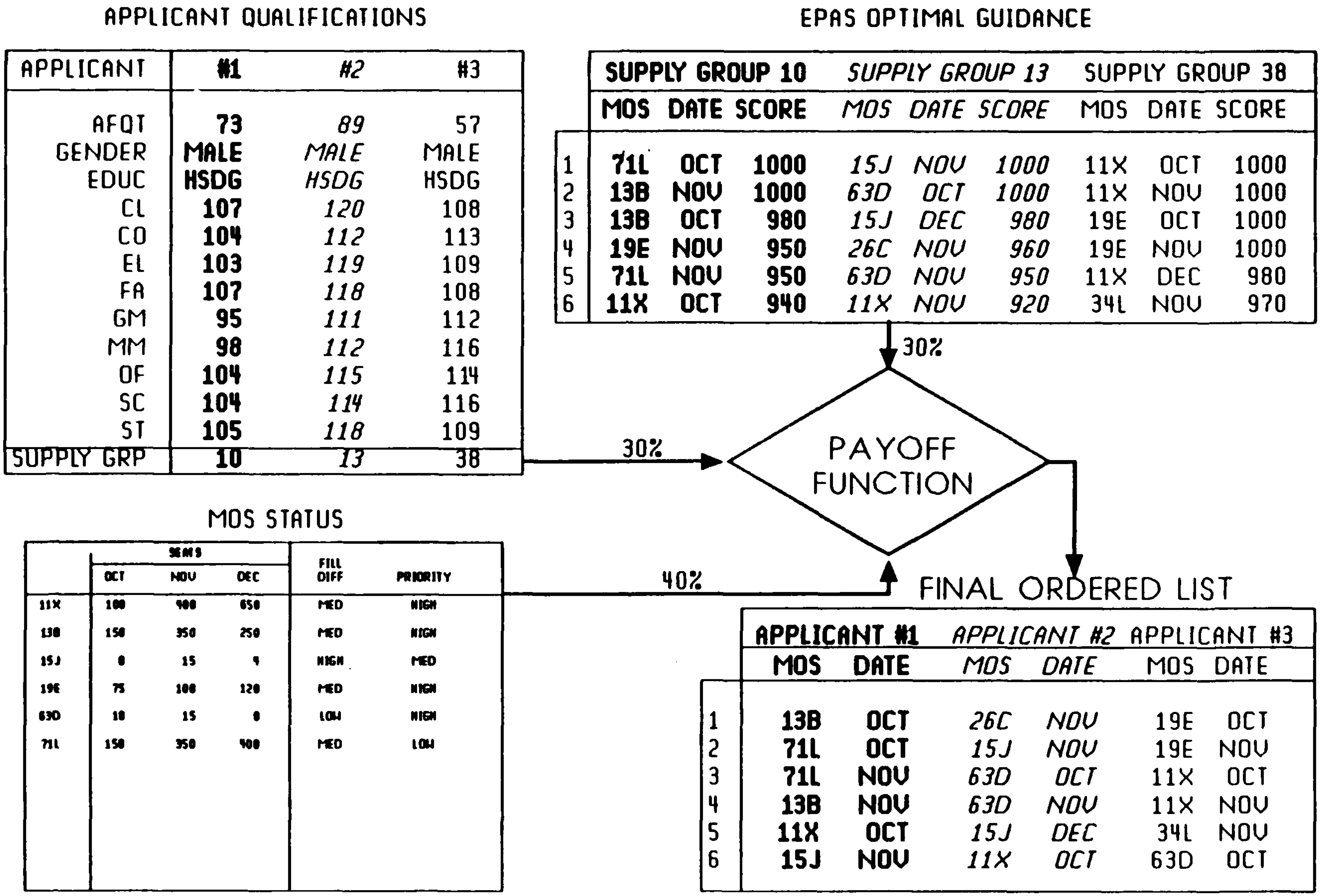


Figure 2.5. EPAS Ordered List Generation Flowchart

However, immediate requirements can still override theoretically optimal matches. For example, applicant 1 is shown 13B first because of its higher priority, although 71L and 13B were scored equally in the optimal guidance.

The payoff functions used to convert MOS and applicant characteristics into a "score" for each potential applicant-job match also change under EPAS. Figure 2.6 shows how EPAS payoff functions can increase the responsiveness of recommendations with respect to performance, quality goals, and female distribution goals. For example, the present system uses all aptitude area scores in the same way regardless of validity. EPAS scoring reflects the predicted performance of the applicant in each job. The result is that EPAS places applicants into MOS where they perform the best.

EPAS is designed to accommodate the fact that actual assignments are sequential. The decision algorithm used by the applicant classification subsystem is similar to that used by other services. It evaluates specific MOS for each applicant when he or she reports to sign an enlistment contract. The alternatives are evaluated according to three kinds of factors: applicant characteristics, job characteristics, and optimization guidance.

The planning information is combined with detailed data on applicant and job characteristics so that each recruit can be evaluated against the actual training seats available to him or her. Table 2.8 lists the factors present in the EPAS algorithm. There are two significant differences between the EPAS algorithm and the present Hierarchy weights. First of all, applicant characteristics that predict performance are included. Second, an interaction term between MOS and applicant is included. This interaction term, supplied through a look-up table, is used to provide guidance from the planning system that assures that the many goals and constraints on the distributional aspects of the assignment are met while performance is improved.

The frequent updates of the forecasts and planning guidance assure that the recommendations remain on track with policy objectives. Since the DEP holds over three months of recruits, there is time to correct errors in forecasts.

Major changes in recruit supply or training plans could require separate analysis. In fact, one additional benefit of EPAS is the capability to perform simulation analyses of the impact of changes in the recruiting environment quickly. It is possible to investigate how changes in enlistment standards, recruit supply, or training schedules will impact on the performance and costs of the Army's enlisted force.

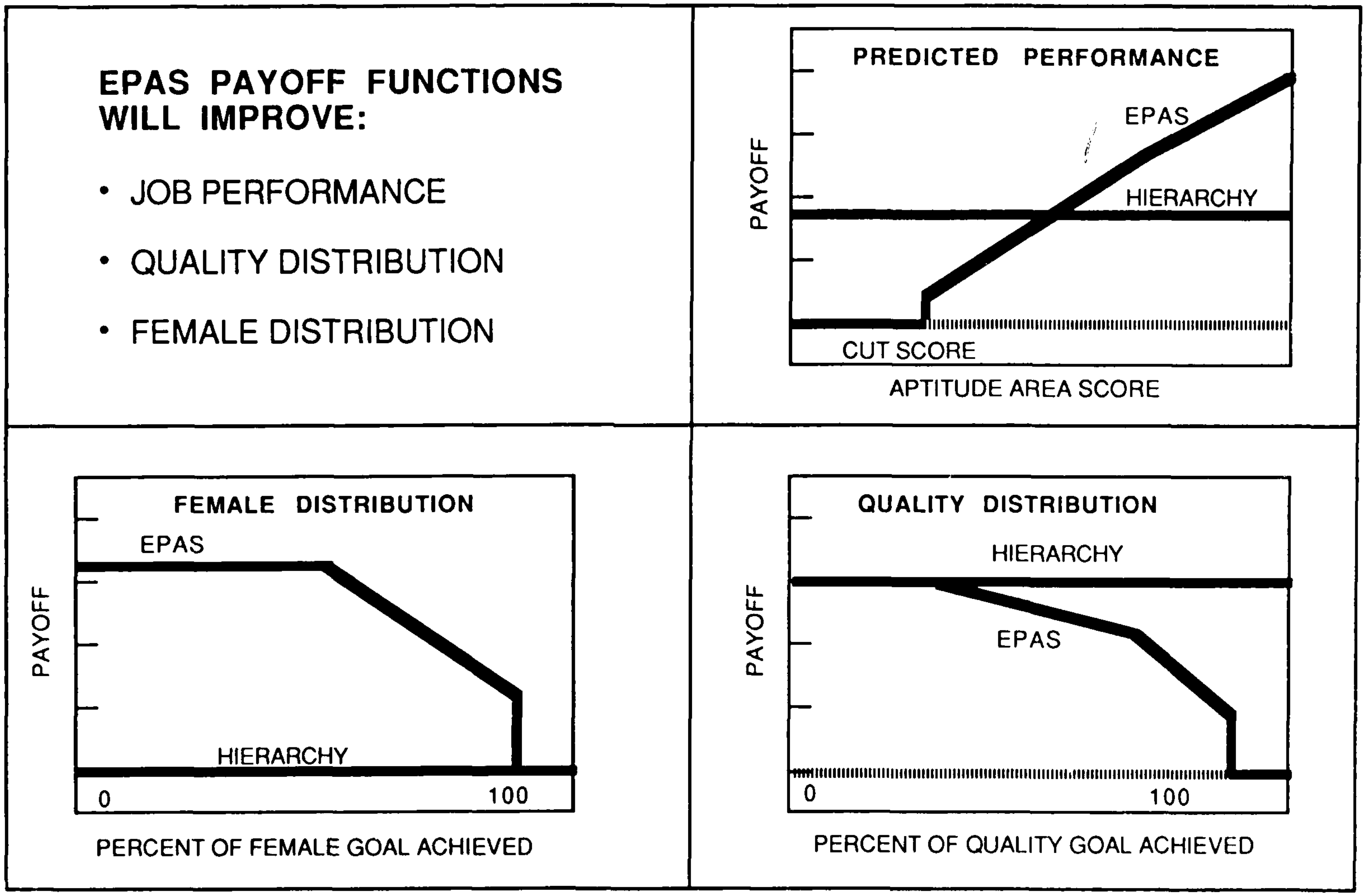


Figure 2.6. Payoff Functions: EPAS vs. Hierarchy

Table 2.8. Factors in MOS Ordering Functions

	Hierarchy		EPAS	
Factor Type	Factor	Weight	Factor	Weight
MOS Only	Cohort Seats	32%	Cohort Seats	10%
	Requirements	26%	Difficulty of Fill	15%
	Training Seats	17%	Time to Fill	10%
	Class Seats	15%		
	Priority	10%	Priority	10%
Applicant Only	Aptitude Area SCR	0	Pred. Performance	10%
	AFQT	0	Quality Goals	5%
	SEX	0	Female Goals	5%
	MEPSCAT	0	Pred. Attrition	10%
MOS and Applicant	----	--	Optimal Guidance	30%

This approach provides a number of ancillary advantages. The impact of assignment decisions on important management indicators can be calculated. The Army can assess how assignment strategies will affect job performance, attrition, MOS fill, and the composition of the DEP. Also, EPAS can be used to simulate alternative policies and environments. The impact of changing requirements, recruit supply, or enlistment standards can be evaluated prior to their occurrence.

Previous analysis by Armor et al. (1982) indicated that it would be beneficial to the Army to raise job standards. Fernandez and Garfinkle (1985) found that Army job performance could be improved and attrition lowered through optimal job allocation. However, they did not know if it would be feasible to achieve these gains operationally because their simulations did not take into account the kinds of restrictions on filling jobs that occur in the real world, nor did they account for the fact that applicants are permitted to reject matches they find undesirable. Nevertheless, they estimated that the benefits from

improved job assignment would require over 200 million dollars to achieve through higher salaries in 1980 dollars.

A number of simulations have been performed of the EPAS system as part of its research and development. The first set of simulations (Schmitz and McWhite, 1986) was primarily to determine whether the overall concept was feasible. The results included a variety of alternative supply and demand scenarios, as well as a comparison to a simpler model that did not include the planning module.

The results indicated that the EPAS design could achieve the distributional objectives required by the Army while improving job performance and reducing attrition. No attempt was made to cost out the gains in job performance. However, the attrition savings were estimated to range from 27 to 41 million dollars, depending upon the scenario. While this was a partial evaluation, it clearly indicated that the benefits of implementation would substantially outweigh an estimated annual operating cost of 650 thousand dollars.

A pilot evaluation of the gains from the EPAS system was performed by Schmitz and Nord (1987). They used both economic substitution costs and utility theory to compare both the benefits of changing the aptitude area composites and the assignment algorithm. They found: (1) the benefits of changing the assignment system to EPAS were in the hundreds of millions of dollars under any benefit-cost framework; (2) the introduction of new composites that reduced the intercorrelation among the aptitude area scores was worthless using the present assignment scheme; and (3) changing the composites under a scenario where EPAS was in place would generate 25 million dollars in additional benefits.

The next chapter details the gains resulting from simultaneous changes in job entry standards and assignment procedures using EPAS.

E. DISCUSSION

The research and development of improvements to the Army's allocation system indicates that substantial benefits from increased job performance are available at negligible cost to the organization. These gains can be achieved using existing information on job performance without disrupting current personnel policies or procedures.

A number of improvements are still possible using the present allocation system. New predictors of attrition and other important dimensions of job performance are being

developed by Project A, ARI's major research effort to validate and expand the current predictors of performance. To the degree this research effort is successful in improving potential allocation efficiency it can be useful to the Army's person-job matching system.

Data on differential job utility would also be useful to incorporate in a new allocation system. Project A has assessed the relative value of different performance levels in all entry level MOS (Nord & White, 1988). This information can be used to weigh alternative assignments in terms of the Army's payoff, rather than simply in terms of performance gains.

Another source of data that is important for allocation is the cost of recruiting and training recruits. One would like to take into account the cost of attrition in different MOS, not simply the probability of attrition. The Army Manpower Cost Model (Horne, 1987) can be used to provide such data.

Finally, the discussion of testing and assignment policy usually is concerned with a single decision--the matching of new recruits with training slots. While this is a critical decision, the operational system eventually is confined to consideration of meeting job quotas with the use of numerical standards. Alternative recruiting, selection, assignment, and retention policies may be equally or more important.

In order to develop a way to deal with a broader array of issues with regard to testing and human resource management it is necessary to estimate realistically and accurately the benefits and costs of alternative policies. Policy makers can then make rational choices on the same bottom-line basis as other organizational interventions are made. The next chapter addresses these issues.

CHAPTER 3. ESTIMATING PERFORMANCE AND UTILITY EFFECTS OF ALTERNATIVE SELECTION AND CLASSIFICATION POLICIES

Roy D. Nord and Edward J. Schmitz

A. PURPOSE AND ORGANIZATION

The analysis described in this chapter has three purposes:

First, to measure the potential gains in Army enlisted soldier performance in each of the Army's nine job families that can be achieved through simultaneous changes in job entry standards (cut scores) and allocation procedures.

Second, to obtain realistic estimates of the costs and benefits of these performance gains in dollar terms.

Third, to place these estimates on a continuum anchored at one end by the performance levels that would obtain if the entire process of selection, classification and job allocation were purely random, and at the other by the performance levels that would occur if the Army were free to place every selected applicant in the job yielding the highest expected performance. Our purpose here is to allow a variety of policies, varying in terms of practical feasibility as well as cost to be compared to each other in relative terms.

The most fundamental requirement for such an effort is that it provide decision-makers with *realistic* information that can be used to make rational choices with respect to the allocation of *scarce resources* among alternative strategies for improving organizational productivity.

The italicized words in this statement are critical: The predictions of the analysis with respect to costs and benefits of the policies being assessed must be "realistic" in the sense that they are not dependent on assumptions about individual and organizational behavior that are unlikely to hold true in practice. Theoretical soundness and logical consistency are necessary, but not sufficient for a meaningful result. Secondly, the analysis must accommodate the fact that resources are scarce--that is, external constraints

impose limits on the range of feasible actions. It is not sufficient to show that a given investment of resources will produce a net positive return. The decisionmaker must also be able to determine that the gains from that investment are equal to or greater than those that would accrue from alternative uses of the same resources--i.e., the "opportunity costs" of the policy must be considered.

Taken together, the need for realism and the need to consider opportunity costs imply that a utility analysis should be context-specific. The assumptions relied on, factors included in the analysis, and the metrics used to calibrate costs and benefits will depend on the organizational context within which the analysis will be used, the set of alternatives being compared, and the objectives that matter to the decisionmaker using the analysis.

The most important difference between the work we describe in this chapter and previous utility studies such as those described in Zeidner and Johnson (1989) is that this analysis has been carried out in response to a demand from military policymakers for more and better information on the comparative merits of alternative manpower policies. The issues under consideration include the proper role of job standards in the military selection process, the potential payoffs to improvements in performance measurement and prediction, and the impact of implementing a new Enlisted Personnel Allocation System (EPAS).

All of these areas involve complex changes in current policy, as well as significant implications for the cost of manning the force. Previous studies have focused on demonstrating the usefulness of utility analysis as a decision tool, but have had little if any impact on actual decisions. Our analysis is an application of utility analysis within the decision process. This difference has two consequences: it results in a focus on the inter-relationships among selection, classification and allocation that is not evident in previous work, and it requires more careful attention to the labor-market consequences of selection policies.

Thus, while we rely heavily on the work of previous researchers in this area, our analysis extends previous work in several key respects:

1. The simultaneous consideration of selection, classification and assignment, with particular attention to the interdependencies among the three processes.
2. The use of empirically based simulations, rather than theoretically derived relationships to obtain estimates of performance gains under alternative policies.

3. The incorporation of labor market considerations into the estimation of selection costs.
4. The transformation of manpower requirements ("quotas") into output-based, rather than input-based units (productive man-months, rather than qualified accessions).
5. The use of expected duration of service and expected training costs as endogenous variables in the calculation of productivity gains.
6. The use of "opportunity costs" as well as net present value as utility metrics.
7. The comparison of the effects of using a "full-least-squares" predictor of performance for selection and classification with those obtained using a single composite predictor of job performance.

The chapter is organized as follows: Section B develops a conceptual model of the problem we are addressing. The purpose of this section is twofold: first to provide a framework that can be used to structure the remainder of the discussion, and second to place the problem within the context of economic theory and clarify some of the assumptions and simplifications used in the the analysis. Section C describes the selection and classification policies simulated, the methods employed to simulate the policies and the data used in the simulations. Section D discusses the distributions of predictors and predicted performance across jobs produced by the simulations. Section E describes the methods and assumptions used in the cost-benefit analysis. The cost-benefit results are presented in Section F. The final section of the chapter provides a brief discussion of the implications as well as limitations of this research.

B. A CONCEPTUAL MODEL OF PERFORMANCE ALLOCATION

In this section we develop a model of the acquisition and allocation of human capital within the context of classical economic production theory. We begin by proposing a simple model of optimal allocation of resources and showing how the issues of personnel selection and allocation can be integrated into such a model. We focus particular attention on the usual assumption in the personnel psychology literature of constant marginal costs of labor.

The basic model is extremely simple, but nevertheless useful as a structure within which we can explicate the assumptions, limitations and rationale that underlies our analysis. Note: The following notational conventions will be used in this chapter: Italicized arabic characters (x, Z) are used to represent scalar variables; bold lower case

arabic characters (**y**) represent vectors of variables; and bold uper case is used to represent matrices. Parameters are represented with italicized Greek characters (ν). Finally, functions over variables are denoted with standard lower case arabic characters (f(x)).

Assume that the Army's objective is to maximize total output (Q) subject to a budget constraint (c^*). For the purposes of this discussion, we shall assume that Q is a scalar quantity that can be measured in dollars. Output could be multidimensional and measured in physical units, but such a specification would considerably complicate the the model without adding substantially to its usefulness for our purposes. To further simplify, we also assume that total output is a simple additive function of a vector of outputs $q = [q_1, q_2, ..., qm]$ from a set of m Army jobs (or job clusters)--specifically Q = **qw**, where **w** is an mxl vector of weights reflecting the relative value of job output. This assumption implies that the contribution of output from each job to total output is independent of the mix of output levels across jobs. While this is, in general, an unreasonable assumption, it is likely to be approximately true as long as the mix of output levels across jobs is not drastically changed. [For a discussion of recent research of variations in peformance value across jobs, see Sadacca, White, Campbell, DiFazio, and Schultz (1989).]

For each job (j), we assume that output is a function of the level of job performance (z_j), as well as other inputs such as equipment, materiel, etc. (x_j) allocated to that job. (Note: For purposes of this exposition, we shall pretend that "job performance" can be increased only by "purchasing" higher quality labor. In practice, of course, job performance is itself a function of many variables.)

We assume further that (a) job performance is measured on an interval scale, (b) the level of labor quality in a job can be adequately represented by the mean level of job performance across the workers in that job, and (c) that the *quantity* of labor (i.e., the number of workers) in each job is fixed. This specification is obviously an oversimplification, but the model could be easily expanded to accommodate variations in the number of workers and a non-linear aggregation of individual levels of performance to the job level.

The cost of producing output is the quantity of each input times its average cost. The classical model of production assumes that, at the level of a single firm, input prices are independent of the quantity of inputs used--that is, that price, average cost, and

marginal cost are all constant. This assumption follows from the condition that where there are a large number of firms each firm's demand for inputs is small relative to aggregate demand for those inputs; thus changes in a firm's demand are too small to cause changes in the input's price. For our purposes, this assumption must be relaxed. The military's demand for high-quality recruits represents a significant proportion of the total youth population, and previous research (e.g., Daula and Smith, 1986; Fernandez and Garfinkle, 1985) strongly indicates that the marginal cost of high quality recruits increases with Army demand. Since the cost of obtaining willing applicants with high levels of predicted performance is central to the problem of estimating the costs of increased selectivity in recruiting, it is important that the phenomenon of increasing marginal costs be incorporated in our model. We therefore allow for the possibility that cost per unit increase in the average performance level may depend on the level used, by specifying the cost function for performance as c(Z) where Z is simply the average level of performance across all jobs, i.e.,

$$Z = \frac{1}{N}\sum_{j=1}^{m} n_j z_j \quad .$$

Note that this specification implies that the cost function is the same for all jobs. If job performance (labor quality) can be differentially predicted by job, then this may be a bad assumption, but we shall address this later. To simplify notation, we assume constant marginal costs ($\mathbf{p_x}$) for other inputs.

Given these assumptions, the objective of maximizing output subject to a budget constraint can be expressed as

$$\text{Maximize } Q = \sum_{j=1}^{m} f_j\,(z_j\,,\,x_j\,) \tag{3.1a}$$

$$\text{Subject to } \sum_{j=1}^{m} c(Z)z_j + P_x x_j \le c^* \tag{3.1b}$$

Combining the two equations using the method of LaGrange yields the following expression to be maximized:

$$\sum_{j=1}^{m} f_j\,(z_j\,x_j\,) - \lambda\left[\sum_{j=1}^{m}\Big(c(Z)z_j + P_x x_j\Big) - c^*\right], \quad , \tag{3.2}$$

where λ is a Lagrangian multiplier which can be interpreted as the "shadow price" of the budget constraint--that is, λ is the increase in output that could be obtained if the budget were increased by one unit.

For the remainder of this discussion, we impose the additional assumption that the production functions (f_j) have the same general form in all jobs. In our analysis we assume, not only that the production functions have the same general form, but that the parameters on job performance are also the same in all jobs. (This occurs because we assume that this parameter is a function of wages, and our measure of wages is the same for all jobs.) For a discussion of the consequences of this assumption, see Nord and White (1988).

The marginal product of an input is the change in output produced by a small (e.g., one unit) change in that input, holding all other inputs constant. Thus, the marginal product of high quality labor in job j is

$$MP(z_j) = \frac{\partial Q}{\partial z_j} = \frac{\partial f(.)}{\partial z_j} \quad , \tag{3.3}$$

where ∂ is the partial derivative operator.

The most common assumption in the personnel psychology literature, which we shall also use in our net present value analysis, is that the change in output resulting from a change in job performance (i.e., $MP(z_j)$) is a linear function of the standard deviation change in performance. If the values of z_j are distributed normally, this implies that $MP(z_j) = \alpha_j(z_j - \mu)/\sigma$, where α_j is the linear parameter on the standard deviation (usually assumed to be a proportion of mean wages in job j), and μ and σ are the mean and standard deviation, respectively, of z_j. Note that this implies that marginal product is increasing for $-\infty \leq z_j < \mu$ and decreasing for $\mu < z_j \leq \infty$. It also means that, in the dimension of performance, the production function has the shape of the normal distribution function. While such a specification for a production function is unusual in the economic literature, it is consistent with the observation in the operations research literature that organizational "personnel response" functions generally display increasing marginal returns over some range, followed by decreasing returns at higher levels of mean performance (Mason and Flamholtz, 1978). This function has the interesting property that, in the presence of a linearly increasing cost function, the returns to a small increase in performance may be negative, while those to a larger increase are positive.

The marginal cost of an input is the change in total cost that will result from a one unit increase in the quantity of the input used. In this model, the marginal costs of all inputs other than job performance are constant. For job performance, marginal cost is dependent on the level of performance, that is

$$MC(z_j) = \frac{\partial c(Z)}{\partial z_j} \quad . \tag{3.4}$$

A common assumption in the personnel psychology literature is that $MC(z_j)$ is constant--simply a function of testing costs. In this case, our analysis departs from the usual assumptions. We assume that the marginal cost of attracting high quality applicants increases with the rejection ratio--that is, as an organization becomes more selective, it must pay the price exacted by the fact that more highly qualified applicants have attractive alternative opportunities. The rate at which marginal cost increases with the selection ratio will depend on (a) the extent to which the selection instrument measures characteristics that are valued by competitors in the labor market (i.e., the extent to which it measures *general* as opposed to *firm-specific* human capital); and (b) the extent to which it measures *either* general or firm specific human capital more accurately than do instruments available to the competition. Both specificity and accuracy confer competitive advantages, and thus lower the marginal cost of obtaining high quality applicants, though it is likely to be the case that the "edge" gained via accuracy (validity) will be a temporary one, since other organizations can presumably develop or purchase the same degree of accuracy over time. The advantage of specificity (differential prediction), however, is more permanent. The effect of improving the measurement of firm- (or job-) specific human capital is to restrict the pool of competitors to those firms or organizations that value the same specific skill, and thus to lower marginal costs on a permanent basis. (Note, however, that this effect will be limited by the "opportunity wage" of the potential applicant--that is by the market value of skills that are more widely valued--and thus by the degree to which job-specific predictions are intercorrelated.)

The **optimal mix** of inputs can be determined from the conditions for a maximum of equation 3.2. These conditions simply state that, when the equation is at a maximum, one of two conditions must hold:

(a) all of its first derivatives must be equal to zero, or

(b) $\lambda > 0$, implying that equation is constrained from further increases by one of its bounds.[3]

If condition (b) holds, then the maximum will occur at the boundary. Note that, if the conventional assumptions (both MC and MP constant) are imposed, then this must be the case. If condition (a) holds, the following relationships must hold at optimality:

$$\frac{\partial f(.) / \partial z_j}{\partial f(.) / x_j} = \frac{\partial c(Z) / \partial z_j}{\partial x} \quad , \text{ for all j } . \tag{3.5}$$

That is, the ratio of the marginal gain from an increase in performance to the marginal gain from increasing some other input must be equal to the ratio of the marginal costs of those increases. This is a key equation for determining whether an increase in the selection ratio, given an inflexible budget constraint, is justified. An increase in the selection ratio justified only if, for some element of x_j, the left side of equation 3.5 is greater than unity and the right side is less than unity. This situation is illustrated (using equipment as the x element) in Figure 3.1. The curve QQ' is a *production isoquant* representing the set of alternative combinations of equipment and soldier quality that can be used to produce a fixed level of output, Q. The curve PP' is a parallel production isoquant for a higher level of output, P. The curve CC' is a "budget isoquant" representing the alternative mixes of equipment and soldier quality that can be obtained with a fixed budget, C. The concave shape of the production isoquants implies that a higher level of output can be obtained with a combination of equipment and soldier quality than could be obtained using either input exclusively. The fact that they are parallel implies that the relationship between the two inputs and output is independent of the output level. The slope of the production isoquants is given by the left-hand side of equation 3.5. The convex shape of the budget isoquant, on the other hand, derives from our assumption that the marginal cost of labor quality is increasing. The slope of the budget isoquant is the negative of the right-hand side of equation 3.5. Given increasing marginal costs for labor quality, the numerator of this expression will increase as z_j increases. Since the denominator is a constant, the slope of CC' must become increasingly negative as the level of soldier quality increases.

---

[3] More precisely, if a solution to the first-order conditions exists, then that solution is either a maximum or a minimum of the equation. Second order conditions must be checked to determine which. If the maximum occurs outside the feasible range of one or more of the arguments (i.e., if $\lambda > 0$), then a more comprehensive set of conditions, the Kuhn-Tucker conditions, must be examined. (See, e.g., Varian, 1978.)

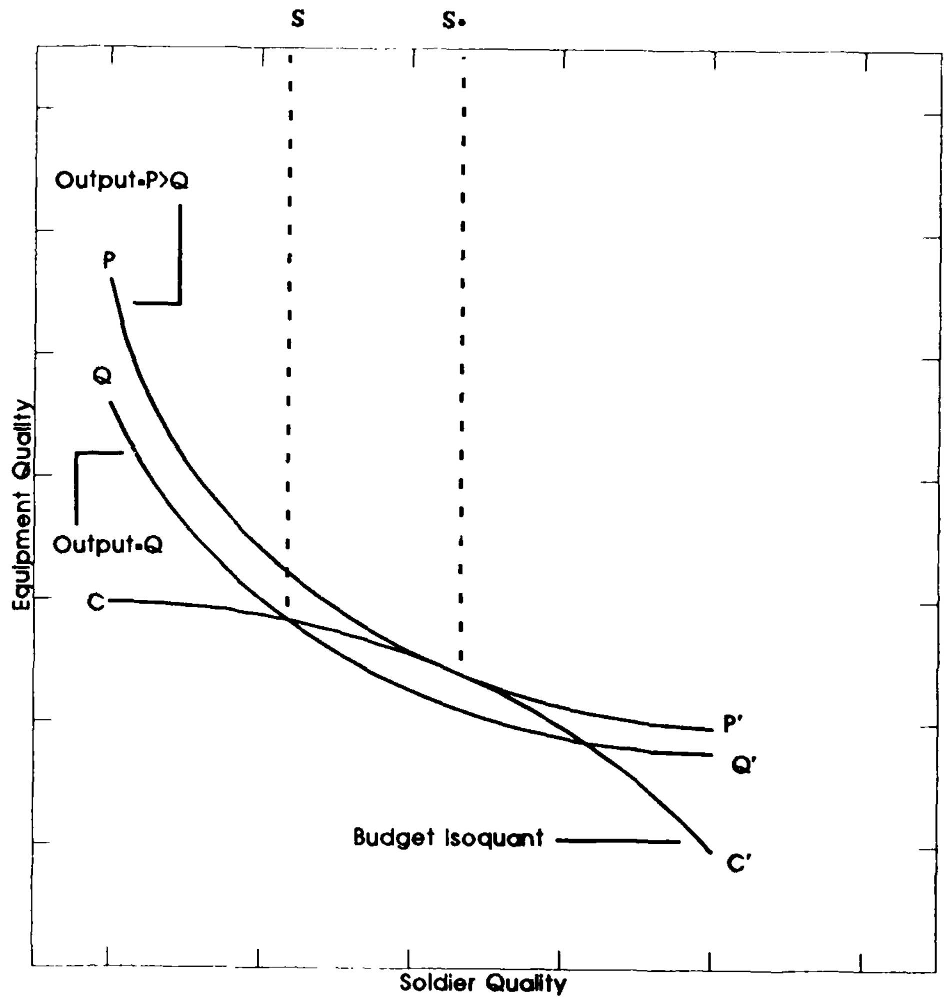


Figure 3.1. Optimal Selection with Increasing Marginal Manpower Costs and a Budget Constraint

(If marginal costs for equipment are also increasing, the shape is simply more exaggerated.) If the organization is currently operating at the point designated by S, we can see that a decrease in the proportion of the budget spent on equipment and a corresponding increase in expenditures on soldier quality is needed to move to the optimal point S^* (allowing input to increase from Q to P *with no increase in the budget*). Assuming that the most efficient way to increase soldier quality is through increased selectivity, this would imply that increasing the selection ratio is cost effective in this example.

For the case of optimal allocation of manpower across jobs (as opposed to resources across inputs), the relevant requirement is that

$$\frac{\partial f(.) / \partial z_j}{\partial f(.) / z_m} = \frac{\partial c(Z) / \partial z_j}{\partial c(Z) / z_m} \quad \text{, for all j, m} \tag{3.6}$$

where both j and m index jobs. This requirement simply states that the ratio of payoffs in different jobs must equal the ratio of marginal costs. However, since we are assuming that the marginal cost curve is the same for all jobs, equation 3.6 reduces to

$$\frac{\partial f(.) / \partial z_j}{\partial f(.) / z_m} = 1 \tag{3.7}$$

implying

$$\frac{\partial f(.)}{\partial z_j} = \frac{\partial f(.)}{\partial z_m} \quad . \tag{3.8}$$

Figure 3.2 illustrates this situation. In this case, the X and Y axes represent two different jobs. Here *AA'*, *BB'*, and *CC'C"* are again "budget isoquants" representing the set of attainable mean predicted performance levels that can be obtained in the two jobs under three different assumptions about the degree of correlation between predicted performance in the two jobs. If predicted performance levels are perfectly correlated, the line *AA'*, with slope equal to the negative of the ratio of the predictor validities (v_1/v_2), is the relevant frontier,if performance is perfectly *uncorrelated*, *CC'C"* is the result. (In this case, the ratio of the validities is the ratio of the y to x intercepts.) Finally, if predicted performance levels are correlated at a level between 0 and 1, the curve *BB'* is the relevant one. The slope of this curve at a given point (x_1^*, x_2^*), representing a particular pair of mean performance **predictor** scores in jobs 1 and 2, respectively, is equal to

$$-\frac{v_1}{v_2}\frac{\partial\phi(x_1, x_2)}{\partial x_1 \partial x_2}\bigg|_{x_1^*, x_2^*},$$

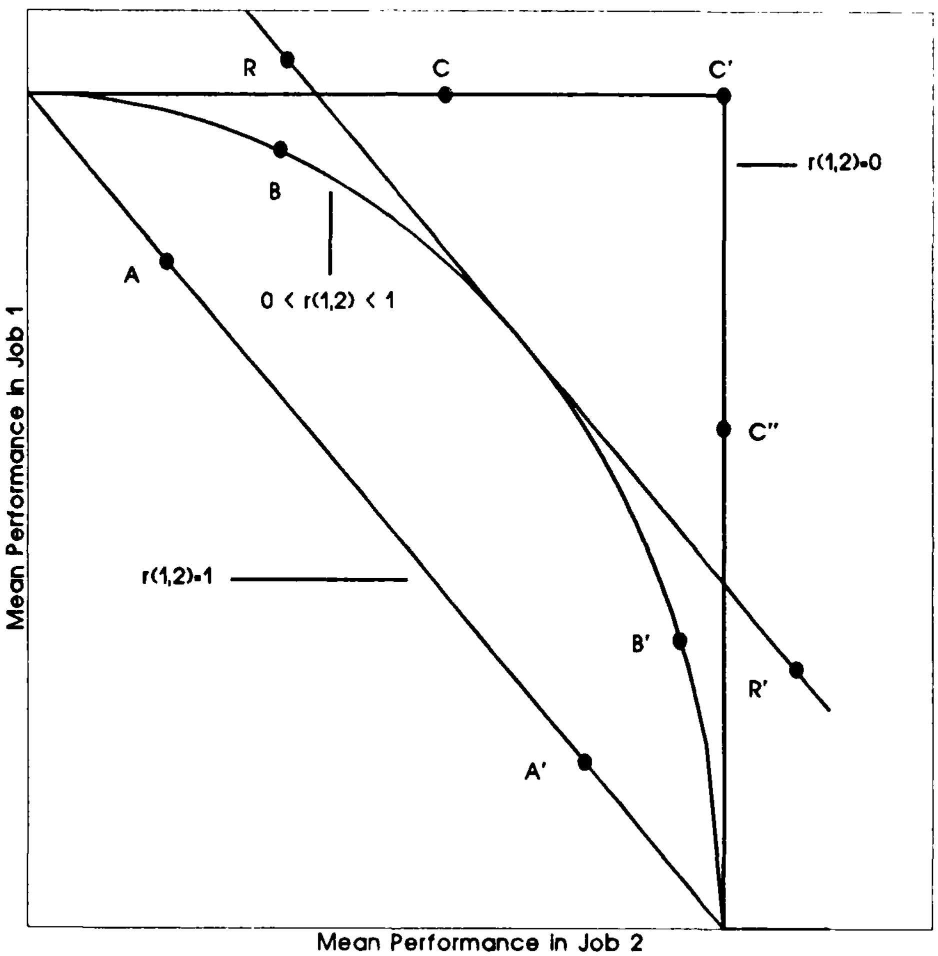


Figure 3.2. Production Possibility Frontiers for Two Jobs for Different Degrees of Predictor Intercorrelation

where ϕ is the normal density function (assuming the predictor scores are distributed as a bivariate normal).

The line RR' is again a production isoquant, the set of all possible combinations of performance in the two jobs that will yield some fixed quantity of total output. Its slope is the negative of the ratio of the marginal products of performance in jobs 1 and 2. The tangency between RR' and BB' is the allocation yielding the highest possible performance level from the available population. Note that the point of tangency will always occur at C' in the orthogonal case; and at one of the intercepts in the perfectly correlated case. The problem of optimal allocation is simple in either of these instances--for the orthogonal case the optimal policy is simply to assign each applicant to the job where he has the highest predictor score. In the perfectly correlated case, the rule is only slightly more complex: if the ratio of marginal products is smaller than the ratio of validities, assign the best applicants to the job with the highest validity (pure hierarchical classification), otherwise assign the best applicants to the job with the highest marginal product. It is only in the case of imperfect correlation (or non-constant marginal productivity) that a search for the point of tangency (i.e., optimization) is necessary.

C. SIMULATING SELECTION AND ASSIGNMENT POLICIES

1. Policies

The manpower procurement policies simulated for this analysis are designed to illustrate the effects of two kinds of changes in the current Army policies--first, changes in the minimum aptitude area scores required in each MOS, and second, changes in the way aptitude area scores are used to make job assignment decisions. The simulations also vary with respect to the kinds of operational constraints on selection and classification they incorporate. This variation provides an opportunity to explore, not only the theoretical effects of changes in selection and classification procedures, but also the potential gains from relaxing or modifying current operational constraints.

A total of thirty-three different policies were analyzed--eleven different job assignment procedures were first simulated under 1984 entry standards, then under the assumption that those standards were raised by five points for all Army jobs (Plus5), and finally under the assumption of a ten point across-the-board increase in standards (Plus10). All thirty policies were simulated using the same random sample of 4377 accessions from 1984 Army enlistments. In addition, to verify the stability of both performance predictions

and cost-benefit estimates, nine of the policies were simulated using two different "synthetic" applicant pools. A brief description of the assignment policies and the methods used to simulate them follows:

Current: The Army's current selection and classification system is described in some detail in the preceding chapter. This policy was not actually "simulated." Instead, the actual assignments under 1984 standards were used to calculate a baseline set of average performance scores for each of 36 job clusters.

The selection of an appropriate sample to simulate the policies involving increased job standards required some assumptions as to how such policies would be implemented under the current selection system. We initially considered simply eliminating from the sample those individuals who would fail to meet the higher standard, but this approach resulted in unrealistically high rejection ratios under the Plus5 and Plus10 policies. (Eighteen percent of actual accessions would have been eliminated under the Plus5 alternative, and 36 percent under the Plus10 option.) Examination of the "rejected" pool revealed that a large proportion of the pool would have qualified for several jobs under the increased standards, but were rejected because they were marginally qualified for their 1984 assignment. It seemed reasonable to assume that, had higher standards been in effect, at least some proportion of these individuals would have been accepted and assigned to a job for which they were qualified. Whether or not this would occur under the current assignment system would depend on the availability of class seats within the time "window" open to the potential "rejectee." In light of these considerations, we used the following procedure to determine whether or not an individual in the base sample would be rejected under each increased standard: First, the base sample was sorted by month in which the contract was signed. Within each month, the scores of potential rejectees were examined to identify job clusters for which the individual was qualified under the new standard. (The sequence of clusters examined varied, depending on the job originally assigned.) If a feasible alternative was found, the set of individuals originally assigned to that job in the same month was searched to identify a candidate who was qualified to take the place of the potential rejectee. If such an individual could be found, the two job assignments were reversed, and the "rejectee" was retained in the sample. If no qualified candidate for a "trade" in any feasible job cluster could be found, the "rejectee" was eliminated from the sample. This approach produced rejection rates (relative to the base sample) of 4 percent and 10 percent for the Plus5 and Plus10 policies, respectively.

We have taken this approach, rather than attempting to devise a true simulation of the current system because (a) the complexities of the current system defy accurate representation in a simulation model, and (b) because this approach produces an optimistic estimate of the capabilities of the current system. Since one of the objectives of this analysis is to determine whether changes in that system are warranted, this strategy amounts to a "hedge" against mistaken rejection of the null hypothesis that no change is needed.

Random: No performance information is used for job assignment. Individuals meeting the relevant standard (Current, Plus5, or Plus10) in 1984 sample are randomly reassigned.

EPAS: The Enlisted Personnel Allocation System described in Chapter 2 is used to assign the sample. Job standards, quality goals, gender restrictions, cohort unit targets, Delayed Entry Program (DEP) policies and class seat constraints are enforced. Applicants are sequentially assigned in the order of their contract signature date. Optimization is used to guide the sequential assignments. The objective function in the EPAS assignments is specified to maximize AA score in the assigned job.

For the simulation under current job standards, EPAS is used directly to assign the base sample. However, because EPAS is currently undergoing modifications prior to implementation, we were unable to simulate either the changes in job standards or the effect of using different metrics in the objective function. We have attempted to approximate the effect of changes in job standards under EPAS by (a) eliminating individuals classed as "rejected" under the current system from the original set of EPAS assignments, and then using a procedure similar to that described above to reclassify any remaining infeasible assignments. (Note: There were very few of these reclassifications, since the EPAS optimization tends to produce relatively few marginal assignments.) This *ad hoc* approach can be remedied by conducting actual simulations with EPAS in the future.

Constrained "Top Down" Assignment: Decision rule assignment--individuals are assigned in order of contract date ("door date"). The assignment pool contains the same individuals as "CURRENT" under each standard. Assignment is to the job family in which they have the highest AA score if (a) the quota for that job is not yet filled, and (b) the AA score meets or exceeds the minimum for that job. If the job is filled, assignment is determined by the individual's next highest score, and so on. If all jobs for which the individual is qualified are filled, the individual is "rejected." Mean performance

in each job is estimated as the mean of the individuals assigned. (This is equivalent to an assumption that empty slots would be filled with individuals having the same mean performance as those assigned.) The number of "rejectees" to be replaced is recorded and used in the cost benefit calculations of recruiting cost under this policy.

Unconstrained "Top Down" Assignment: Same as above except that quotas are ignored. Individuals are simply assigned to job with highest score. (Note: In the cost-benefit analysis, this alternative is treated as if there were no quotas.)

Batch Optimizations: The remaining six allocation policies simulated all use a "batch" optimization to make the assignment decisions. All of these alternatives used a capacitated network assignment algorithm to maximize an objective function subject to supply and demand constraints and to the minimum standards under each policy, but did not enforce the other policy constraints used in EPAS. The policies differed with respect to (a) whether hierarchical classification was used (i.e., whether the objective function maximized predictor scores or predicted performance), (b) whether or not the optimization was allowed to *select* as well as assign applicants, and finally, with respect to the predictors used to predict performance. The optimization problem in each case was of the general form

$$\text{Max} \sum_{i=1}^{N} c_{ij}\, x_{ij}$$

$$\text{Subject to:} \quad \sum_{j=1}^{J} x_{ij} \leq 1 \quad \text{for all i}$$

$$\sum_{i=1}^{N} x_{ij} \geq d_j \quad \text{for all j}$$

where

N is the number of individuals being assigned.

J is the number of job families.

c_{ij} is the Aptitude Area score of individual i in job family j.

x_{ij} = 1 if individual i is assigned to job j, 0 otherwise. (x_{ij} is constrained to be zero if individual i's AA score is below the cut score for job j).

d_j is the demand (quota) in job j.

The five batch optimization allocations are as follows:

Optimization on AA Score (Classification Only) (OPTAACL): This policy used batch-mode optimal assignment to maximize average AA score in assigned jobs. The optimization was allowed to reassign individuals "selected" by the current system under each standard, but *not* allowed to optimize selection from the original applicant pool.

Optimization on AA Score (Selection and Classification) (OPTAASC): This allocation was identical to OPTAACL, with the exception that, under the PLUS5 and PLUS10 selection standards, the optimization was allowed to optimally *select* applicants from the same sample under current standards. The relative rejection ratios (4 percent and 10 percent) obtained for the current system under the Plus5 and Plus10 alternatives were used to constrain selections. Job quotas were set at the levels obtained under the Current system at each selection standard. Note that, while this provides some indication of the potential gains from simultaneous selection and classification, the gains are understated because the optimization was forced to select from the previously restricted sample.

Optimization on Single Composite Predicted Performance (Classification only) (OPTPRFCL): This allocation method was the same as OPTAACL except that c_{ij} is defined as the product of the aptitude area composite for the assigned job and its validity. The validities used are shown in Table 3.6.

Optimization on Single Composite Predicted Performance (Selection and Classification) (OPTPRFSC): This alternative bears the same resemblance to OPTPRFCL as does OPTAASC to OPTAACL. That is, individuals are both selected and classified so as to maximize the predicted performance.

Optimal Assignment (Full Least Squares Prediction) (OPTFLS): This option is the same as OPTPRFSC except that predicted performance is defined as a weighted sum of all composites, where the weights are the least squares coefficients. The procdures used to obtain these weights are discussed in the section on performance measures below.

Optimal Assignment (Full Least Squares Prediction, Quality Goals Enforced) (OPTFLSQG): This option is the same as OPTFLS except that the optimization is constrained to allocate a minimum percentage of AFQT category I-IIIA recruits to each job family. The "quality goals" used were those actually in effect in 1986, scaled to reflect the lower overall proportion of CAT I-IIIA accessions in our sample.

Table 3.1 provides a summary of all these policies.

Table 3.1. Summary of Simulation Scenarios

SCENARIO IDENTIFIER	SELECTION STANDARD	OPTIMAL SELECTION	CLASSIFICATION CRITERION	ALLOCATION METHOD	SIMULATION METHOD
RANDOM0	None	No	None	Random	n.a.
RANDOM1	Current	No	None	Random	n.a.
RANDOM5	Plus 5	No	None	Random	n.a.
RANDOM10	Plus 10	No	None	Random	n.a.
CURRENT1	Current	No	AA score	Current	n.a.
CURRENT5	Plus 5	No	AA score	Current	Extrapolated from actual
CURRENT10	Plus 10	No	AA score	Current	Extrapolated from actual
EPAS1	Current	No	AA score	Optimal/Seq	EPAS
EPAS5	Plus 5	No	AA score	Optimal/Seq	Extrapolated from EPAS
EPAS10	Plus 10	No	AA score	Optimal/Seq	Extrapolated from EPAS
OPTAACL1	Current	No	AA score	Optimal	Network opt
OPTAACL5	Plus 5	No	AA score	Optimal	Network opt
OPTAACL10	Plus 10	No	AA score	Optimal	Network opt
OPTAASC1	Current	No	AA score	Optimal	Network opt
OPTAASC5	Plus 5	Yes	AA score	Optimal	Network opt
OPTAASC10	Plus 10	Yes	AA score	Optimal	Network opt
OPTPRFCL1	Current	No	Val*AA score	Optimal	Network opt
OPTPRFCL5	Plus 5	No	Val*AA score	Optimal	Network opt
OPTPRFCL10	Plus 10	No	Val*AA score	Optimal	Network opt
OPTPRFSC1	Current	No	Val*AA score	Optimal	Network opt
OPTPRFSC5	Plus 5	Yes	Val*AA score	Optimal	Network opt
OPTPRFSC10	Plus 10	Yes	Val*AA score	Optimal	Network opt
MAXAACON1	Current	No	AA score	Sequential	Rule: best available job
MAXAACON5	Plus 5	No	AA score	Sequential	Rule: best available job
MAXAACON10	Plus 10	No	AA score	Sequential	Rule: best available job
MAXAAFREE1	Current	No	AA score	Sequential	Rule: best job (no quota)
MAXAAFREE5	Plus 5	No	AA score	Sequential	Rule: best job (no quota)
MAXAAFREE10	Plus 10	No	AA score	Sequential	Rule: best job (no quota)
OPTFLS1	Current	No	Σ wt*AA score	Optimal	Network opt
OPTFLS5	Plus 5	Yes	Σ wt*AA score	Optimal	Network opt
OPTFLS10	Plus 10	Yes	Σ wt*AA score	Optimal	Network opt
OPTFLSQG1	Current	No	Σ wt*AA score	Optimal	Network opt with Qual Goals
OPTFLSQG5	Plus 5	Yes	Σ wt*AA score	Optimal	Network opt with Qual Goals
OPTFLSQG10	Plus 10	Yes	Σ wt*AA score	Optimal	Network opt with Qual Goals

2. Data

a. Empirical Sample

A random sample of 4377 individuals was used to simulate the assignment of recruits to 36 clusters of jobs. The job clusters are differentiated on the basis of Aptitude Area composite and the minimum score on the composite required. Table 3.2 shows the distribution of quotas across the clusters for the samples used in our simulations and for 1984 Army accessions.

Table 3.3 provides summary statistics on the distribution of predictor scores in the youth population (McLaughlin et al, 1984). The intercorrelations among the predictors range from 0.67 to 0.97, with an average intercorrelation of about 0.85.

The observed intercorrelations among predictor scores for the empirical simulation samples under each selection standard are shown in Table 3.4. As one would expect, the average intercorrelation declines as the sample becomes more restricted. The mean intercorrelation among predictors for the population selected under current standards was about 0.81. This drops to 0.78 when standards are raised by 5 points, and to 0.74 when standards are increased 10 points. The standard deviation of the predictors also declines in each case. The predictors each have a standard deviation of 20 in the youth population. The mean standard deviation in the 1984 accessions sample is 12.36, declining to 11.57 when standards were raised 5 points, and to 10.44 under the Plus10 scenario.

b. Synthetic Samples

The use of a sample of actual accessions to simulate the effect of alternative selection and classification policies has both advantages and disadvantages. The advantage of this approach is that the sample (at least in the case of current selection standards) has been selected by a "real" as opposed to hypothetical selection process. As was noted in Chapter 2, the Army's selection process relies not only on the uniform application of a known standard (AFQT score), but also on other criteria that apply differentially across the test score distribution. Furthermore, the distribution is censored in its upper regions as a result of self selection among potential applicants with high test scores. Thus the conventional assumption that the selected population is simply a left-truncated normal distribution with known (population) parameters is unrealistic. [Murphy (1989) partially addressed this issue, but treated it essentially as a problem of truncation from the right,

Table 3.2. Job Demands: Actual vs. Empirical Simulation Sample

APTITUDE AREA	CUT SCORE	PLUS0 SAMPLE	PLUS5 SAMPLE	PLUS10 SAMPLE	1984 ACCESSIONS
ALL	ALL	4377	4200	3939	120281
CL	90	128	100	78	3691
CL	95	329	291	250	11227
CL	100	1	1	1	88
CL	105	3	2	2	133
CL	110	10	9	9	291
CO	90	951	921	860	26978
CO	100	5	5	5	228
EL	90	285	273	258	4897
EL	95	119	116	114	3065
EL	100	33	33	33	628
EL	105	13	13	13	375
EL	110	29	29	29	768
EL	115	1	1	1	74
EL	120	11	11	11	318
FA	85	158	129	109	4942
FA	100	64	64	63	1603
GM	85	5	4	0	203
GM	90	117	115	111	3970
GM	95	34	34	31	1053
GM	100	121	121	119	3992
GM	105	3	3	3	158
MM	85	18	18	17	49
MM	90	303	293	280	8175
MM	100	48	48	48	1568
MM	105	228	228	228	5382
OF	90	226	213	183	7190
OF	100	99	99	99	3149
OF	105	10	10	10	424
SC	95	4	4	3	192
SC	100	10	10	9	329
ST	85	4	3	2	151
ST	90	274	267	232	7528
ST	95	364	363	359	7681
ST	100	330	330	330	7782
ST	105	36	36	36	1436
ST	110	3	3	3	63

Table 3.3. Predictor Correlations in the Youth Population

	CL	CO	EL	FA	GM	MM	OF	SC
CL	--							
CO	.80	--						
EL	.73	.89	--					
FA	.84	.94	.91	--				
GM	.67	.90	.96	.84	--			
MM	.75	.93	.88	.84	.93	--		
OF	.83	.94	.88	.88	.91	.97	--	
SC	.96	.91	.82	.87	.82	.88	.94	--
ST	.76	.89	.96	.90	.94	.87	.92	.84

Note: All scores have a mean of 100 and a standard deviation of 20.
Source: McLaughlin and Rossmeissl, 1985

Table 3.4. Summary Statistics and Predictor Correlations for Simulation Samples: Random Sample of 1984 Army Accessions

a. Current Selection Standards

MEAN	104.5	107.6	105.1	105.4	105.9	107.7	107.3	107.3	104.5
STD	12.0	12.3	12.3	12.2	13.2	12.3	11.3	12.7	12.8
CL	--								
CO	0.71	--							
EL	0.89	0.77	--						
FA	0.87	0.87	0.82	--					
GM	0.72	0.80	0.92	0.68	--				
MM	0.53	0.86	0.74	0.65	0.86	--			
OF	0.67	0.89	0.75	0.72	0.84	0.93	--		
SC	0.77	0.91	0.84	0.76	0.88	0.86	0.93	--	
ST	0.86	0.77	0.91	0.80	0.87	0.71	0.83	0.88	--

SAMPLE N: 4377

b. Selection Standards Raised Five Points

MEAN	105.4	108.6	106.1	106.2	107.0	108.6	108.2	108.4	105.6
STD	11.5	11.4	11.7	11.7	12.3	11.5	10.4	11.7	12.1
CL	--								
CO	0.67	--							
EL	0.89	0.74	--						
FA	0.86	0.85	0.81	--					
GM	0.69	0.77	0.90	0.64	--				
MM	0.47	0.84	0.69	0.60	0.84	--			
OF	0.62	0.87	0.70	0.67	0.82	0.92	--		
SC	0.74	0.90	0.81	0.73	0.86	0.84	0.91	--	
ST	0.85	0.73	0.90	0.78	0.86	0.67	0.79	0.85	--

SAMPLE N: 4200

c. Selection Standards Raised Ten Points

MEAN	106.5	109.8	107.3	107.3	108.3	109.8	109.4	109.8	107.0
STD	10.7	10.1	10.7	10.7	11.1	10.3	9.2	10.2	11.0
CL	--								
CO	0.60	--							
EL	0.87	0.68	--						
FA	0.84	0.82	0.78	--					
GM	0.65	0.71	0.89	0.59	--				
MM	0.38	0.81	0.63	0.53	0.81	--			
OF	0.55	0.84	0.64	0.62	0.77	0.91	--		
SC	0.69	0.88	0.76	0.68	0.83	0.81	0.89	--	
ST	0.83	0.67	0.89	0.76	0.83	0.59	0.73	0.82	--

SAMPLE N: 3939

rather than one of censoring.] On the other hand, the distribution of characteristics in the empirical sample is also partially determined by transient factors (such as economic conditions) that limit the generalizability of findings based exclusively on a single such sample.

To explore the robustness of our results under differing assumptions about the selection mechanism we generated two synthetic populations: (a) a synthetic population with predictor scores with the same means and standard deviations observed in our empirical sample and an intercorrelation matrix with the same *expected* value as the observed matrix; and (b) a synthetic sample with means, standard deviations and expected intercorrelations equal to those in the youth population.

The synthetic 1984 accession sample was constructed in the following manner. First, nine independent normal deviates, each with N = 4500, expected mean of 0, and expected standard deviation of 1, were created. We designate this 4500×9 matrix $\mathbf{X}$. Second, we computed a factorization $\mathbf{F}$ of the intercorrelations among AA scores in the accession population, such that $\mathbf{F}\mathbf{F}^T = \mathbf{R}$. (Note: The full 9×9 matrix $\mathbf{R}$ was effectively singular, so $\mathbf{F}$ was computed as a 9×8 matrix, using a generalized inverse of $\mathbf{R}$. The method for computing this inverse is described in the section on the full least squares predictor below.) Third, the set of scores for the synthetic population used in the simulations was computed as $\mathbf{Y} = \mathbf{X}\mathbf{F}^T$. The matrix of intercorrelations among the pseudo-scores in $\mathbf{Y}$, designated by $\overline{\mathbf{R}}$, has expected value $\mathbf{R}$. Finally, this set of scores was transformed to have the observed vector of means and standard deviations.

The "youth population" synthetic sample was generated in the same way, but the matrix $\mathbf{R}$ contained the population intercorrelations in shown in Table 3.3, and the N was 8000 to allow for selection. The population analogues for the Current, Plus5, and Plus10 selection standards were created by truncating the sample at the percentile equivalent to each selection ratio. The effect of using this sample is that the standard deviations of the resulting populations are much larger than in the previous two samples. For example, the standard deviation of the synthetic sample based on the current selection standards was 16.3, compared to 12.4 for the observed population. This is to be expected, considering the lack of any censoring at the high end of the distribution for this sample.

Table 3.5 shows the grand mean across all AA scores, mean standard deviation and mean intercorrelation for for the empirical samples and both synthetic populations.

Table 3.5. Comparative Statistics for Empirical and Synthetic Samples

SELECTION STANDARD	STATISTIC	EMPIRICAL SAMPLE	SYNTHETIC SAMPLE (1980 Youth Pop)	SYNTHETIC SAMPLE (1984 Accessions)
CURRENT	SAMPLE N	4377	3998	3993
	MEAN AA SCORE	106.2	107.0	106.1
	AA SCORE STD	12.4	12.2	16.3
	MEAN CORRELATION	0.79	0.78	0.79
PLUS5	SAMPLE N	4200	3943	3897
	MEAN AA SCORE	107.1	107.3	106.6
	AA SCORE STD	11.6	12.0	16.1
	MEAN CORRELATION	0.76	0.77	0.79
PLUS10	SAMPLE N	3939	3780	3603
	MEAN AA SCORE	108.3	108.1	108.2
	AA SCORE STD	10.4	11.5	15.6
	MEAN CORRELATION	0.72	0.75	0.77

Tables 3.A1 and 3.A2 in the Appendix show the full intercorrelation matrices for both synthetic samples.

3. Performance Measures

a. Single-Composite Validity Estimates

Table 3.6 shows the average job performance validities for nine occupational clusters or job families (all validities include corrections for range restriction). The aptitude area composites are constructed from tests on the Armed Services Vocational Aptitude Battery (ASVAB). The tests used for each composite are also shown in Table 3.6.

Maier and Grafton (1981) validated ASVAB version 8, 9, and 10 composites against Army Skill Qualification Tests (SQTs) for five job families, final training grades in three other job families, and against final course grades in one job family. In all, 35 different Military Occupational Occupational Specialties (MOS) were validated, employing samples ranging in size from 100 to over 2000 in each MOS. Table 3.6 shows that the mean validity across all jobs in the Maier and Grafton study is 0.60.

McLaughlin, Rossmeissl, Wise, Brant, and Wang (1984) also validated ASVAB 8/9/10 composites against Army SQTs for 46 MOS, employing samples ranging from 1,300 to 16,000 in each MOS. Table 12.6 shows that the mean validity in this study is 0.47, a considerably lower estimate than the 0.60 reported by Maier and Grafton.

Table 3.6. Average Corrected Job Performance Validities of Aptitude Area Composites Used for Assignment to Army Job Families in 1984

Job Family	Aptitude Area Composite	Tests Comprising Composite	Mean Validity 1981[a]	1984[b]	1987[c]	1988[d]
Clerical/Administrative	CL	(VE+NO+CS)	.53	.49	.60	.59
Combat	CO	(AR+CS+AS+MC)	.56	.44	.54	.55
Electronics Repair	EL	(GS+AR+MK+EI)	.59	.45	.72	.65
Field Artillery	FA	(AR+CS+AS+MC)	.63	.45	.39	.55
General Maintenance	GM	(GS+AS+MK+EI)	.76	.40	.54	.60
Mechanical Maintenance	MM	(NO+AS+MC+EI)	.52	.45	.62	.55
Operators/Food	OF	(VE+NO+AS+MC)	.61	.50	.61	.60
Surveillance/Communication	SC	(VE+NO+CS+AS)	.55	.47	.55	.55
Skilled Technical	ST	(GS+VE+MK+MC)	.55	.57	.54	.55

a Maier and Grafton (1981)
b McLaughlin, et al. (1984)
c McHenry (1987); Eaton (1987); Zeidner (1987)
d "Weighted Average" used in present utility analysis (1988)

Description of ASVAB Tests
VE . . . verbal ability (combines paragraph comprehension and word knowledge tests)
NO . . . numerical operations
CS . . . coding speed
AR . . . arithmetic reasoning
AS . . . auto and shop information
MC . . . mechanical comprehension
GS . . . general science
MK . . math knowledge
EI . . . electronics information

The McLaughlin, et al. study used a criterion-referenced SQT criterion that lacked the discriminability and variance associated with norm-referenced tests (Zeidner, 1987).

McHenry (1987) reported Army ASVAB validities against very carefully defined and measured job criteria, including hands-on and job knowledge tests, to minimize problems of reliability and criterion contamination and deficiency. The McHenry study included nine Army MOS, and used samples ranging from 400 to 600 per MOS. Average validity against a job-specific core technical skills criterion was found to be 0.63. When ASVAB composites were combined with other cognitive and non-cognitive predictors, average validity increased to 0.67 against the same criterion. (These results were reported in detail by Zeidner, 1987.) Eaton (1987) reported these same results along with validities for ten additional MOS, using school knowledge and proficiency ratings as criteria.

The single-composite validities used in the present analysis are shown in the last column of Table 3.6. These estimates were obtained by combining the results of the three previous studies using weights based on the results of previous weighted Army validities

as well as the number of MOS in each job family covered by each of the three studies. The resulting "weighted average" falls within the range of previous estimates for each job family, and lies close to the unweighted average across all previous studies. The main effect of this approach is to dampen the large and often inconsistent variation in validities across job families.

To provide a basis for comparison across services, and to show the relationship between training and job performance validities, Table 3.7 shows training validities, corrected for restriction in range, of ASVAB composites by military service (Hunter, Crosson, and Friedman, 1986). Training criteria, as typically defined, are final course grades. The four-job-family structure used in Hunter, et al. analysis includes a wide sampling of occupational specialties. Across all services, the analysis includes 190 jobs and a sample size of 103,700. The overall mean validity is 0.58. Zeidner (1987) notes that the validities used for the Army sample in this analysis were based on the criterion-referenced SQT's reported by McLaughlin, et al. (1984). Zeidner suggests that, for purposes of comparability, it would have been more appropriate to have obtained Army validities using final course grades as criteria, as was done by Maier and Fuchs (1972). The comparable mean validity found in the Maier and Fuchs study is 0.65, based on a sample size of 25,000 in over 100 MOS.

Table 3.7. Average Training Validities of ASVAB Composites for Four Job Families by Military Service[a]

Service	Number of Jobs	Sample	Validity M&C	B&C	E&E	HS&T	Total
Army	55	50,000	.49	.45	.49	.50	.48
Air Force	70	29,700	.70	.74	.77	.74	.74
Navy	31	7,600	.50	.49	.53	.53	.51
Marines	34	16,400	.58	.58	.53	.61	.58
Total	190	103,700	.56	.55	.59	.59	.58

Source: Hunter, Crosson, and Friedman (1985). p. 116.
a. M&C Mechanical and Crafts
B&C Business and Clerical
E&E Electronics and Electrical
HS&T .... Health, Social and Technology

Considering the 1987 finding of average job performance validities of 0.63 for ASVAB composites, and much earlier findings of training validities of the same magnitude, the job performance validities used in the present analysis may be safely considered as conservative.

b. Obtaining "Best-Least-Squares" Performance Predictions

The current Army selection and classification system uses one (occasionally two) of nine Aptitude Area Composites (hereafter AA scores) to predict job performance. The AA score composites are unit-weighted combinations of selected subsets of the ten ASVAB scores. This approach is currently used primarily because of its simplicity. At the time the current selection and classification system was designed, it was essential that the calculations required to determine MOS eligibility be as simple as possible. Given modern computer capabilties, a feasible alternative to this approach would be to use a "full-least squares" (FLS) prediction equation using all ten ASVAB scores to predict performance in each job family.

The FLS predictor equations would be estimated by regressing all predictors against the performance criterion for each job family. This results in a set of predictor weights that are noninteger and frequently negative. To the extent that the information contained in the additional subtests used in each equation is not redundant, and to the extent that the contribution of the subtests to the prediction is not equal, such equations will produce more accurate predictions of true performance than will the single-composite predictors. More importantly, the FLS predictors will provide significantly greater opportunities for *differential* prediction by job family. While the marginal gains in average validity from such an approach (assuming careful development of the unit-weighted composites) are likely to be modest, it is possible that even small gains may be significant given the size of the Army's selection and classification problem. Furthermore, the capacity of EPAS to capitalize on the added potential for classification efficiency offered by the FLS predictor makes this approach worthy of consideration. The potential "gaming" problem produced by the negative weights could be overcome by assuring that the weights are not public knowledge. Finally, much of the earlier theoretical work on classification efficiency assumes that performance predictions are best-least-squares estimates. For these reasons, the present analysis uses an approximation to the true best-least-squares predictions in two ways--as the classification criterion in two of the simulated allocation strategies (OPTFLS and OPTFLSQG); and as the measure of the expected performance produced under all of the simulated policies.

The FLS predictors used in this analysis are an approximation to the true FLS predictors because the regression weights they use are based on the nine AA composites, rather than directly on the ten ASVAB scores. This means that the weights we use are the least squares coefficients that would result from a regression using the tests, subject to a

large set of restrictions on the relative values of the weights. The effect of these restriction is almost certainly a reduction in the goodness of fit of the model, producing in turn an underestimate of the potential gains to be obtained by using a true FLS predictor.

This compromise was necessitated by the fact that, in order to compute the FLS weights without conducting a full-scale validity study, the full matrix of validities of each ASVAB subtest against each job family was required. We were unable to obtain this matrix, and relied instead on the matrix of AA composite validities published in McLaughlin, et al. (1984). This matrix is shown in Table 3.8.

Table 3.8. Original Matrix of AA Composite Validities Against All Job Families

JOB FAMILY	CL	CO	EL	FA	GM	MM	OF	SC	ST
CL	.48	.51	.53	.54	.49	.46	.50	.50	.53
CO	.36	.44	.43	.43	.43	.42	.44	.40	.44
EL	.38	.47	.47	.46	.47	.46	.47	.44	.47
FA	.39	.49	.48	.48	.49	.49	.49	.45	.44
GM	.39	.48	.46	.46	.47	.48	.48	.45	.47
MM	.36	.48	.46	.45	.48	.48	.48	.43	.46
OF	.38	.48	.47	.45	.48	.47	.48	.44	.48
SC	.39	.49	.48	.47	.48	.47	.48	.45	.49
ST	.51	.56	.57	.57	.55	.54	.56	.54	.58

Source: McLaughlin and Rossmeissl (1984)

In order to conform the validities in this matrix to the "weighted average" estimates shown in Table 3.6, the McLaughlin et al. matrix was rescaled to produce the appropriate single-composite validities in its diagonal elements. This was done by simply multiplying each row of the matrix by the ratio of the "weighted average validity" to the original diagonal element. The rescaled validity matrix is shown in Table 3.9.

Given this matrix of validities the FLS weights were calculated as follows:

Let **R** be the (9×9) matrix of correlations among the nine AA composites (shown in Table 12.3), and **V** the (9×9) matrix of correlations between each predictor and job performance in each of the 9 job families shown in Table 3.8. (Note, v_{ij} is the correlation between predictor j and job family i, where i indexes rows and j indexes columns.)

Table 3.9. Adjusted Validity Matrix Used to Generate "Best-Least Squares" Predictor Weights

JOB FAMILY	PREDICTOR CL	CO	EL	FA	GM	MM	OF	SC	ST
CL	**0.59**	0.63	0.65	0.66	0.60	0.57	0.61	0.61	0.65
CO	0.52	**0.64**	0.63	0.63	0.63	0.61	0.64	0.58	0.64
EL	0.44	0.55	**0.55**	0.54	0.55	0.54	0.55	0.51	0.55
FA	0.45	0.56	0.55	**0.55**	0.56	0.56	0.56	0.52	0.50
GM	0.41	0.51	0.49	0.49	**0.50**	0.51	0.51	0.48	0.50
MM	0.41	0.55	0.53	0.52	0.55	**0.55**	0.55	0.49	0.53
OF	0.48	0.60	0.59	0.56	0.60	0.59	0.60	0.55	0.60
SC	0.49	0.62	0.61	0.60	0.61	0.60	0.61	**0.57**	0.62
ST	0.57	0.63	0.64	0.64	0.62	0.61	0.63	0.61	**0.65**

Then **W**, the 9×9 matrix of least squares weights (element w_{ij} being the weight on predictor j for job i), is simply $\mathbf{VR}^{-1}$. Let **S** be a 9×9 matrix with off-diagonal elements equal to 0 and diagonal elements equal to the diagonal elements of the matrix $\mathbf{VR}^{-1}V^T$. Then the diagonal of $\mathbf{S}^{1/2}$ contains the multiple correlation coefficients of the BLS predictors for each job, and the matrix $\mathbf{S}^{-1/2}(\mathbf{VR}^{-1}\mathbf{V}^{\mathrm{T}})\mathbf{S}^{-1/2}$ contains the correlations among predicted performance scores.

Due to the high degree of collinearity in the matrix **R**, it was impossible to obtain $\mathbf{R}^{-1}$ directly. (The errors introduced by rounding the correlations to two significant digits were sufficient to make the matrix singular.) We therefore used a generalized inverse of **R**. The generalized inverse we used was obtained by computing the (9×1) vector e containing the eigenvalues of **R** and the (9×9) matrix **D**, containing the associated eigenvectors. The negative eigenvalue was dropped, along with the associated row of **D**, and the generalized inverse was computed as $(\mathbf{D}\mathrm{diag}(e)\mathbf{D}^{\mathrm{T}})^{-1}$, where "diag" is the matrix operator that transforms a vector into a square matrix with the elements of the vector on the diagonal.

Table 3.10 displays the FLS weights computed by this procedure, along with the multiple correlation coefficients for each job family. The average multiple R is 0.62, a 0.05 increase over the average of 0.57 for the single-composite predictors.

The use of the FLS predictors for both assignment and evaluation in the OPTFLS simulations raises two issues: First, both the validity and correlation matrices used to obtain the FLS weights are estimated with error. These errors are propagated to the weights, and thus to the FLS predictions of performance. The optimization will maximize

Table 3.10. Full Least-Squares Weights (9 composites against 9 job families)

JOB FAMILY	CL	CO	EL	FA	GM	MM	OF	SC	ST	Multiple R
CL	-0.184	-0.019	0.562	0.253	-0.176	-0.519	0.411	0.367	-0.027	0.69
CO	-0.110	0.082	-0.143	0.269	0.140	-0.080	0.256	0.020	0.234	0.66
EL	-0.304	-0.085	0.189	0.276	0.141	-0.259	0.471	0.288	-0.157	0.57
FA	-0.704	-0.949	0.090	1.220	0.872	-0.513	1.075	0.702	-1.249	0.64
GM	0.077	0.094	-0.264	0.159	0.236	0.207	-0.022	-0.110	0.178	0.53
MM	-0.228	-0.082	-0.225	0.383	0.422	-0.026	0.323	0.114	-0.122	0.57
OF	-0.037	0.394	0.058	-0.186	-0.029	0.052	0.023	0.015	0.338	0.62
SC	-0.061	0.489	0.186	-0.184	-0.168	-0.011	0.021	0.030	0.340	0.64
ST	0.377	0.149	-0.107	0.061	0.060	0.308	-0.317	-0.292	0.491	0.67

both the "true" and the error components of the prediction, and the use of the same weights for evaluation treats the error components as gains in true performance, thus overestimating the gains that would be realized under the OPTFLS policies. (Note: This problem is *not* related to "back" validities, since the validities and correlation matrices used to obtain the weights are based on entirely different samples than those used in the simulations.) We suspect that the overestimation produced by this problem is small, and any overestimation that does occur will be at least partially offset by the *underestimation* resulting from the use of composites rather than subtest scores. Nevertheless, we plan to explore the possible effects using model sampling methods in the near future.

The second issue is related to both the errors in the estimated correlations and the near-singularity of the correlation matrix. These two factors combine to produce least-squares weights that are highly unstable. That is, small random variations in the estimated intercorrelations can produce large changes in the weights. While this is a serious problem if one's objective is to obtain reliable estimates of the true weights on each AA composite, it does not necessarily produce unstable *predictions* of performance. We examined the stability of the predictions by creating several sets of weights using small perturbations of the intercorrelation matrices. Each set of weights was used to produce a prediction of performance, and we then examined the correlations among the different predicted values. The resulting correlations all fell between 0.95 and 0.99, thus providing some assurance that the FLS predictions are reasonably stable. However, the planned model sampling exercise will allow us to test this tentative conclusion more rigorously.

4. Simulation Results: Predictor Scores and Predicted Performance

The policies defined above were used to allocate the sample of candidates to jobs under alternative selection standards. We examine the impact of these alternatives on the distribution of mean Aptitude Area scores as well as on the distribution of predicted performance. We report the effects on AA scores for two reasons: First, because predictor scores, as opposed to predicted performance, are the focus of current practice; and second, because the contrasts among between predictor and predicted performance distribution across jobs under different policies provide a useful illustration of the effects of hierarchical classification.

a. Predictor Scores

Table 3.11 shows the mean AA scores across nine job families under each selection and classification policy. The first column shows the results of different strategies under the existing selection standards. Current selection standards raise the average aptitude area score by 6.1 points, or about 0.3 standard deviations over the population mean. The job assignment policies described in Chapter 2 raise the average score by an additional 1.6 points. Under EPAS, the average increase over random assignment is nearly 3.9 points, more than twice the gain over random assignment yielded by the current assignment system.

Table 3.11. Simulation Results. Average Aptitude Area Scores in Assigned Job Family

	SELECTION STANDARDS		
METHOD	CURRENT	PLUS5	PLUS10
RANDOM	106.1	106.8	107.6
CURRENT	107.5	108.7	109.7
EPAS	110.0	110.7	111.9
OPTAACL	113.0	113.8	114.7
OPTAASC	113.0	114.0	115.3
OPTPRFCL	112.8	113.5	114.5
OPTPRFSC	112.8	113.9	115.2
MAXAACON	111.6	112.3	113.2
MAXAAFREE	113.9	114.7	115.9
OPTFLS	108.7	110.2	110.8
OPTFLSQC	109.1	110.3	111.1

Note: Baseline option of random selection and assignment yields average of 100.

The batch optimizations maximizing aptitude area score (OPTAACL and OPTAASC) increase the mean score to 113, or 6.9 points above random assignment. (Note that the results for the classification only and selection and classification alternatives are identical under current standards. This is because the entire available "applicant pool" is being assigned in both cases, thus no gains can be realized from optimal selection.) This allocation yields the maximum attainable average AA score from this population against the existing requirements.

As would be expected, when objective is to maximize single predictor maximization of performance (OPTPRFCL and OPTPRFSC) the resulting allocation yields a somewhat lower average AA score. It is somewhat surprising that the reduction in AA scores is negligible. The batch optimizations policies all increase AA scores by an additional 3 points over the level attained under EPAS.

The results for the two rule-based algorithms also show substantial gains in predictor scores. The constrained algorithm (MAXAACON) provides an average score of 111.64, which is a substantial increase over current policy, but about 1.4 points below the batch optimization results. This policy also exceeds EPAS, because it does not consider many of the additional distribution constraints faced by operational policy. When the requirement of meeting job demands is lifted (MAXAAFREE), the average score increases to 113.9, which is above the optimal assignment, but infeasible. This policy provides an indication of the effect of "non-natural" quotas on potential classification gains.

The OPTFLS policy predictors, while producing smaller increases in AA scores than the other optimizations nevertheless produces a larger gain over random assignment than that produced by the current system. The substantial decline in the average score under this policy, as compared with the maximization of single-predictor performance is a result of both the increased effect of hierarchical classification, and the lower correlation of job-specific composites with the FLS predictor.

Figure 3.3 compares the current distribution of AA scores across the nine Army AA job clusters to that produced by the EPAS, OPTAACL and OPTFLS policies under current selection standards. Note that both EPAS and OPTAACL provide approximately equal or higher averages in each job cluster, while the OPTFLS results show considerably more variability across jobs.

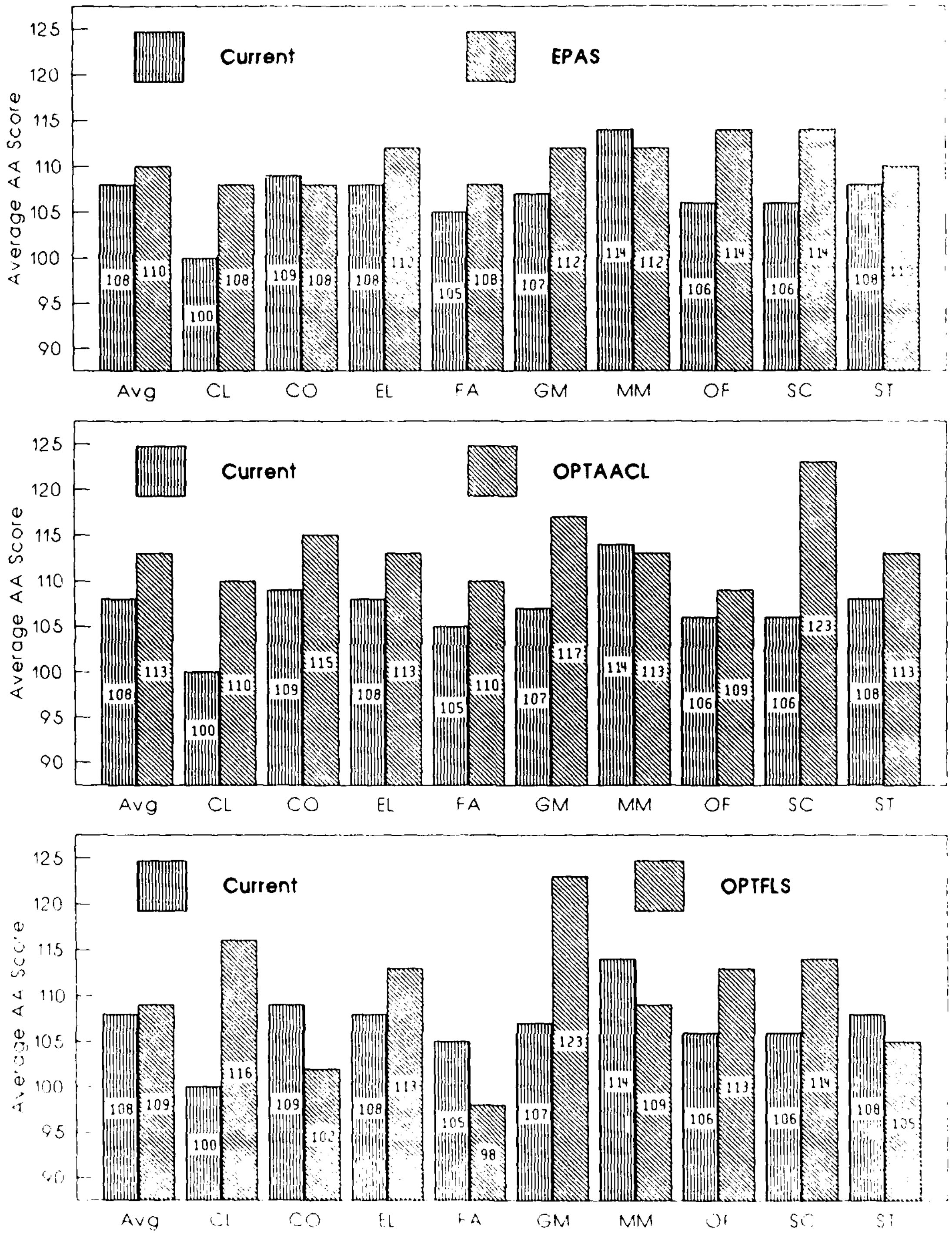


Figure 3.3. Comparison of Mean AA Scores by Job Family Under Current and Alternative Allocation Strategies. Job Standards at Current Levels.

Table 3.11 also provides results for alternative selection standards. The pattern of results that occurs with existing selection standards holds within each of the sets of policies. As expected, increasing the selectivity produces gains in predictor scores. For example, raising the standards by 10 points with the current assignment policy would produce a gain in performance of nearly the same magnitude as using EPAS to assign the existing population.

While scores increase under all policies as selection standards are increased, the gains differ as the allocation policy becomes more efficient at matching applicants with jobs. For example, a 10 point increase in standards increases average performance 1.53 points under random assignment, but 2.25 points when the current system is used. The gains from optimal selection and classification (OPTAASC) are about one point for a five point increase in standards, and 2.27 points from a 10 point increase. The effect of simultaneous selection and classification from this restricted population are small, but noticeable under the Plus10 selection standard, yielding a 0.8 point increase.

b. Predicted Performance

Table 3.12 shows the gains in the FLS prediction of performance under each policy. The gains are calibrated in standard deviations relative to the population mean of 0. For the most part, the performance gains follow a similar pattern to that shown in the predictor scores, with some important exceptions. The gains from EPAS over the current system are proportionately not as great, since EPAS simulations used aptitude area scores as the objective. For the same reason, optimization of single-predictor performance provides greater performance gains than are produced from when AA score is maximized.

Table 3.12. Simulation Results. Average Predicted Performance in Assigned Job Family

	SELECTION STANDARDS		
METHOD	CURRENT	PLUS5	PLUS10
RANDOM	0.189	0.209	0.236
CURRENT	0.197	0.227	0.254
EPAS	0.221	0.242	0.272
OPTAACL	0.236	0.266	0.293
OPTAASC	0.236	0.265	0.303
OPTPRFCL	0.245	0.264	0.297
OPTPRFSC	0.245	0.269	0.312
MAXAACON	0.229	0.250	0.276
MAXAAFREE	0.254	0.279	0.316
OPTFLS	0.340	0.386	0.405
OPTFLSQG	0.330	0.370	0.396

Note: Baseline option of random selection and classification yields average of 0.

The most notable change is the very substantial gain from the OPTFLS policy. This method (OPTFLS) results in increases in predicted performance of nearly 0.1 standard deviations above the OPTPRFSC policies under all three selection standards. The increase over the current system under current selection standards is roughly 1.5 times the gain of the current system over random selection and classification. It is clear that the addition of "real world" constraints like those used in EPAS would curtail these potential gains. As noted earlier, we were not able to directly test the effect of these constraints on the gains provided by the FLS alternative. However, we were able to test the effect of adding one additional constraint--the AFQT quality goals--to the optimization using the FLS predictions. As can be seen in Tables 3.11 and 3.12, the effects of this added set of "real-world" constraints on the optimization results were very small--producing an increase in the average AA score of roughly one point, and a reduction in mean predicted performance of about 0.15 standard deviations. While the reductions due to the other constraints in EPAS may be larger than this, it seems reasonable to expect gains as high as 0.25 to 0.3 standard deviations over random by using FLS predictors in EPAS. This would be a larger gain than that produced by the current selection standards, and would impose far fewer costs.

Figure 3.4 shows the performance distributions across the Aptitude Area clusters under three classification policies and current selection standards. Again the increased variability across jobs under the OPTFLS policy is evident. The combined effect of distributional constraints and the absence of hierarchical classification is evident in the relatively even, but generally lower distribution produced by EPAS. In the cost benefit analysis described below, we will attempt to answer the question of whether these gains are sufficiently large to offset the increased recruiting costs associated with higher standards.

The performance gains from increased selectivity generally fall in the range of 0.05 to 0.07 standard deviations. Note that these gains are generally smaller than the variation across methods within each selection standard. The effect of changes in selection standards across jobs is depicted in Figure 3.5. As might be expected, the increase in standards produces an increase in performance in every job cluster, and the increases tend to be fairly evenly distributed across jobs.

c. Synthetic Sample Results

Table 3.13 compares the results of three simulations using the two synthetic populations described in Section B with those obtained using the empirical sample.

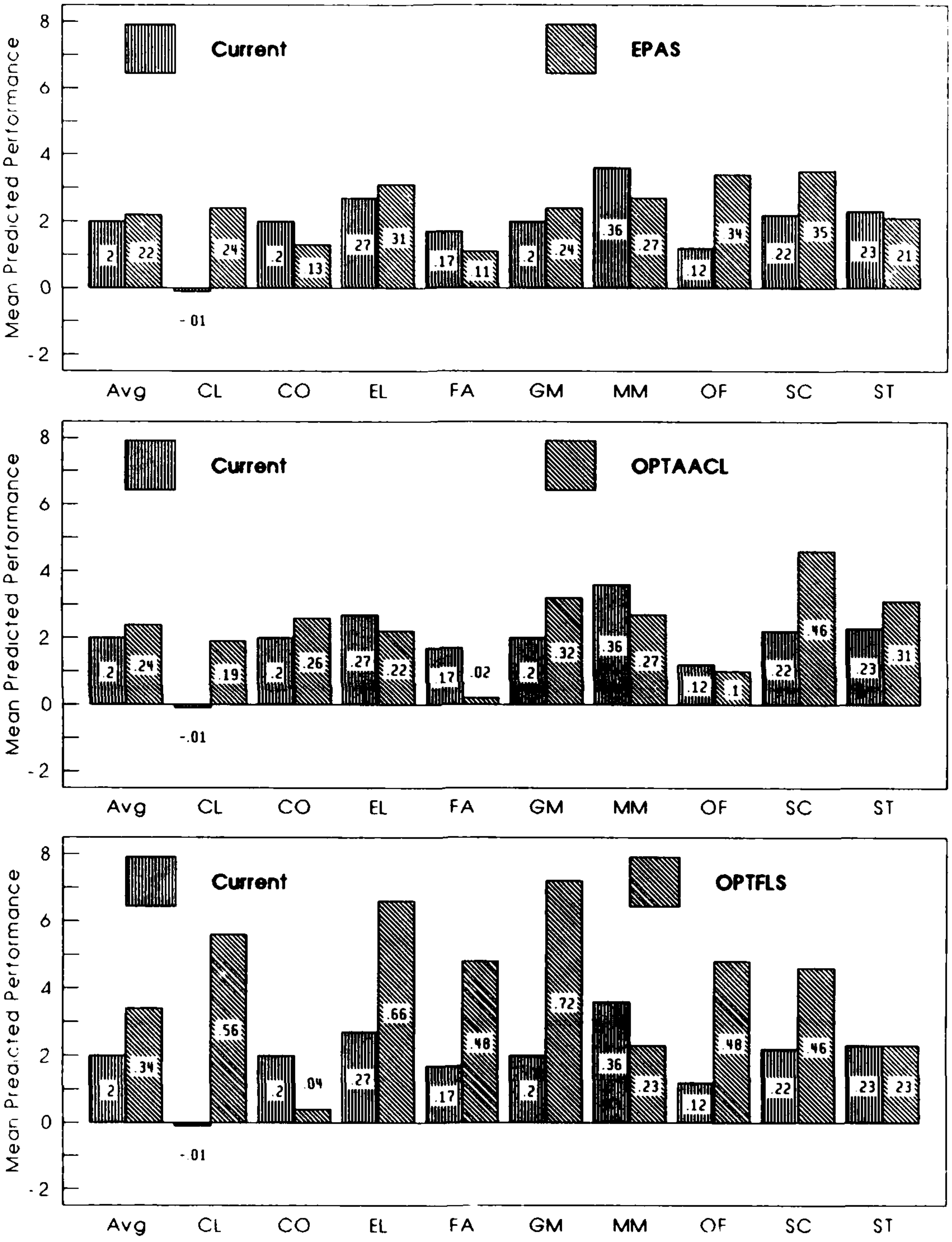


Figure 3.4. Comparison of Mean Predicted Performance by Job Family Under Current and Alternative Assignment Strategies. Cut Scores at Current Levels. (Performance in Standard Units.)

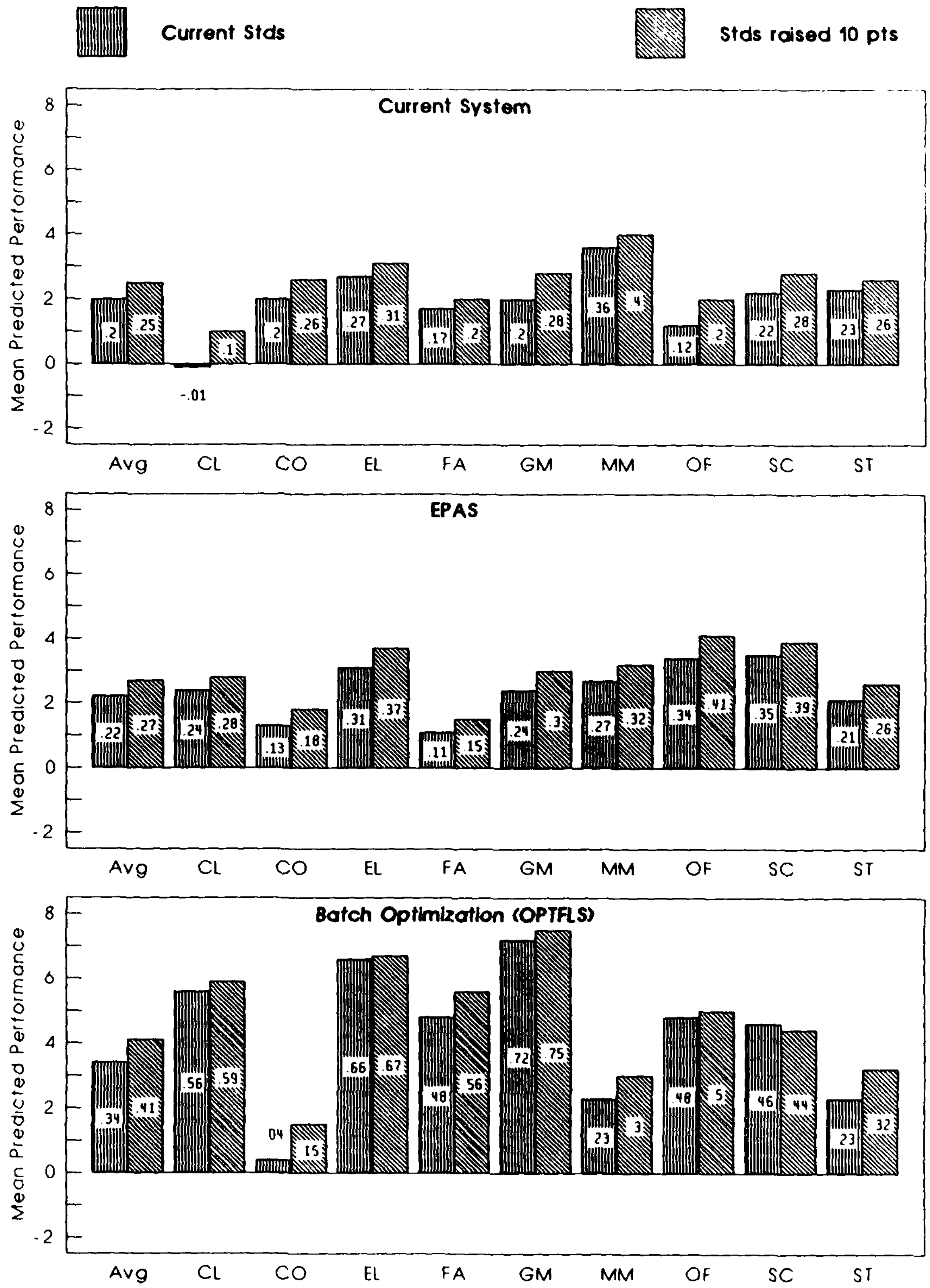


Figure 3.5. Effect of the Selection Ratio on Performance Distribution Under Three Allocation Strategies. (Performance in Standard Units.)

Table 3.13. Mean Predicted Performance and AA Scores by Job Family: Comparison of Synthetic and Empirical Sample Results

SELECT STANDARD	CLASSIFICATION METHOD	SYNTHETIC YOUTH		SYNTHETIC 1984		1984 ACCESSIONS	
		AASCR	PRDPERF	AASCR	PRDPERF	AASCR	PRDPERF
CURRENT	OPTAASC	114.73	0.253	113.95	0.281	113.00	0.236
	OPTPRFSC	114.55	0.263	113.82	0.289	112.82	0.245
	OPTFLS	108.79	0.363	109.63	0.372	108.73	0.340
PLUS5	OPTAASC	115.37	0.273	114.35	0.294	113.76	0.265
	OPTPRFSC	115.72	0.298	114.57	0.311	113.92	0.269
	OPTFLS	109.09	0.377	109.37	0.383	110.23	0.386
PLUS10	OPTAASC	117.19	0.327	115.17	0.322	115.34	0.303
	OPTPRFSC	117.78	0.362	115.74	0.347	115.17	0.312
	OPTFLS	109.08	0.394	109.16	0.397	110.80	0.405

In general the results are remarkably similar. Both synthetic populations produce slightly higher gains under all alternatives, a result that would be expected, given the censoring in the upper regions of the empirical distribution. The relative magnitudes across both selection standards and classification policies are very consistent, suggesting that the predictions produced from the empirical sample are likely to hold up, at least in relative terms, under a reasonably wide variation in accession populations.

D. ESTIMATING THE NET PRESENT VALUE OF PREDICTED PERFORMANCE CHANGES

In this section we evaluate the performance gains produced above via a benefit-cost model. The performance gains are evaluated using two different methodologies of benefit estimation: one based on the psychological utility theory of output valuation, and an alternative approach using economic opportunity costs. These two very different techniques provide the most robust way possible for generating a consensus as to the benefits of selection and classification testing.

1. Methodology

The net present value (NPV) model for performance valuation is a refinement of approach developed by Brogden (1951) and developed further by many other personnel testing researchers such as Hunter and Schmidt (1982), Cascio (1987b), and Boudreau (1983a). We expand upon the traditional utility model by explicitly taking into account estimates of not only the gain in performance, but the length of time over which the

individual performs, as well as the recruiting costs that result under alternative policies and selection ratio.

The equation we use to calculate the NPV of performance is:

$$NPV = \sum_{i=1}^{N^*} \left[\sum_{t=1}^{39} r_t (1\text{-}ATTPROB_{it})(PERF_i{*}VALUE_t - TRCOST_t) \right] - RECOST_i \quad . \quad (12.9)$$

The terms in this expression have the following meanings:

a. N* is the number of willing applicants that must be attracted, given the selection ratio, to yield the number of qualified accessions needed. The number of accessions is fixed at the level required to produce the same number of productive (i.e., post-training) person-months of service as were obtained from actual 1984 accessions. Thus the number of required accessions depends on the expected attrition rate under the policy being evaluated. That is,

$$N^* \, sf(ATTPROB) = PPM$$

or $$N^* = \frac{PPM}{sf(ATTPROB)}$$

where *PMM* = productive person-months obtained from 1984 accessions,

s = the selection ratio, and

f(ATTPROB) = the expected number of person-months per accession, a function of the probability of attrition in each month.

b. t indexes months (given the mix of two, three and four year terms the average commitment of a recruit is 39 months).

c. rt is a discount factor to deflate the net value of performance t months in the future back to the present. (We assumed a discount rate (net of inflation) of 4%.)

d. $ATTPROB_{it}$ is the estimated probability that individual i will fail to complete at least t months of service. This probability was obtained by estimating separate logistic regressions for each of the 9 aptitude area clusters, AA score and its square as predictors. The relevant coefficient estimates were applied to the assigned AA score of each individual after assignment to obtain the probabilities used in the utility equation.

e. $PERF_i$ is the expected performance of individual i in his assigned job, measured in standard units. The Full Least Squares prediction of performance was used for all scenarios.

f. $VALUE_t$ is the estimated dollar value of a one standard deviation increase in PERF at time t. We used the conservative "rule of thumb" that this value is 40% of salary. Salary was approximated by using "real military compensation" (RMC), adjusted to take into account average promotion rates. Table 3.14 shows the salary profile we used. VALUE was specified to be 0 during training under the assumption that the contribution of trainees to Army output is negligible.

Table 3.14. Salary, Training Cost, and Discounting Assumptions

Month of Service	Expected Grade	Monthly RMC[a]	Monthly Training Cost
1-2	E-1	977	1218+RMC
3-5	E-2	1150	4415+RMC
6-12	E-2	1150	0
12-24	E-3	1230	0
25-39	E-4	1369	0
Nominal Total		43,345	21,193
Discounted To tal		9,776	20,918

a *RMC is "real military compenstion", and takes into account the value of benefits and tax advantages as well as nominal monthly pay. All values are in constant dollars.*

b *The assumed discount rate, net of inflation, is 4%.*

g. $TRCOST_t$ is the average monthly cost of training per trainee in month t. At the time of the analysis we did not have access to reliable MOS-specific training cost data, so the costs used were Army-wide averages for basic and advanced training (MOS-specific costs are now available on the Army's AMCOS system). The basic training cost was applied to the first two months of service, and the advanced cost was applied during months 3-5. Training costs after 5 months of service were assumed to be 0. These costs are also shown in Table 3.14.

h. $RECOST_i$ is the average cost of recruiting an individual in the same ability range as individual i, and is assumed to depend both on the ability level and on the total number of individuals within that category who are recruited--that is, we do not assume constant marginal costs. Three ability ranges were defined: below average (AFQT CAT

IIIB or IV), above average (CAT IIIA), and high (CAT I and II). Marginal recruiting costs for below average individuals were assumed to be constant. (This is equivalent to assuming that these recruits are "demand constrained" over the range covered by our analysis.) For the above average and high categories, average cost per recruit is assumed to rise as the number of recruits processed increases. The rate of increase in average costs was estimated by assuming a current (1984) marginal cost for high-quality recruits of $26,000 (1986 dollars), and a constant pay elasticity of 1 (i.e., that marginal cost increases by one percent for each one percent increase in the number of high quality soldiers recruited). This estimate of marginal cost is generally consistent with estimates in the literature, as is the assumption of an elasticity of 1 [Armor, et al. (1982); Fernandez and Garfinkle (1985); Polich, Dertouzos, and Press (1986)]. In addition to the basic question of how marginal costs change with the number of high quality recruits, there is a second question of how the number of high-quality needed changes as selection standards are increased. We make three different assumptions about this, and provide cost estimates under each assumption.

2. Estimating the Selection Ratios

Selection ratios are used in two ways in this analysis: They are needed to determine the value of N^* under each policy, and they are needed to estimate the gains over random selection and classification provided by current selection practices. The ratios used to obtain N^* under all options other than random selection and classification need only be known relative to the N^* for the current system under current recruiting standards. The "empirical" ratios obtained when the Plus5 and Plus10 base sample were selected are sufficient for this purpose. As indicated in Section B above, these ratios are 0.96 and 0.9 for the Plus5 and Plus10 alternatives. Use of these ratios in equation 3.11 will yield the incremental change in applicants compared to 1984 accessions. These ratios cannot, however, be used to calculate N^* for random selection and classification.

Our estimate of the current effective selection ratio was obtained simply: We calculated the grand mean of all AA scores in the 1984 accession population. Assuming this mean to be distributed normally with mean 0 and standard deviation 20 in the population, allows estimated selection ratio to be estimated as the point at which the normal distribution must be truncated to obtain the observed mean value. We used interpolation of the table published in Brogden (1959, Table 1) to obtain this estimate. The result was an assumed selection ratio of 0.83 for the current system.

3. Estimated Attrition Effects

As equation 3.9 indicates, the expected attrition rate under each policy is a key parameter. The attrition rate affects not only the number of accessions needed to obtain a fixed quantity of "effective man-months" of service, and thus *average* as well as total recruiting costs, but also training and salary costs. Table 3.14 shows these costs. As will be shown in the cost benefit results, the multiple effects of attrition, combined with its large costs, make even small changes in the rate important. While none of our simulations explicitly attempted to minimize attrition costs, previous research (e.g., Schmitz and Manganaris, 1984) has indicated that Aptitude Area scores are inversely related to attrition, and that the strength of the relationship varies across Army MOS. Therefore, to account for the changes in expected attrition rates under the various simulation policies, we estimated a simple logistic regression of AA score (in the assigned job) and its square on first term attrition rates in each of the nine AA clusters. (The sample used for this regression was a 50% random sample of all 1984 accessions into the MOS included in our base sample.) The coefficient estimates from this regression are shown in Table 3.15. These coefficients were applied to the mean assigned AA score in each AA cluster to obtain a predicted attrition rate. (Note: For the two clusters (SC and ST) with no significant coefficients, the rate was assumed to remain at its 1984 level under all policies.)

Table 3.16 shows the predicted changes in numbers of attritions in an accession population of 120,281 relative to that occuring under the current system. The associated changes in training costs under each selection and classification standard are also shown. The changes in recruiting costs due to attrition changes are included in the recruiting cost tables discussed below. Although the changes are quite small in percentage terms, with the largest change (MAXAAFREE, Plus10) amounting to less than 2.5% of accessions, the dollar value of training cost savings is significant. Under current standards, EPAS is projected to provide $13.7 million in training savings. This would rise to $28.7 million under the Plus10 alternative. Even if it assumed that all of the attritees are low quality recruits, the recruiting savings would increase these numbers to $15.2 and $31.8 million. These are likely to be very conservative estimates of the reductions that could be obtained for two reasons: First, because the objective of reducing attrition has not been included in these simulations at all, and previous research (Nelson and Schmitz, 1985) has indicated that it may be possible to obtain significant reductions in attrition while retaining nearly all of the gains in performance when attrition is added to the objective function. Second, the

Table 3.15. Logistic Regression Coefficient Estimates Used to Predict Attrition Effects

JOB FAMILY	INTERCEPT	AA SCORE	SQUARED AA SCORE
CL	-9.878	0.171	-0.00082
	(3.453)	(0.065)	(0.00030)
CO	-6.000	0.104	-0.00051
	(1.571)	(0.029)	(0.00013)
EL	-6.330	0.118	-0.00060
	(3.324)	(0.060)	(0.00026)
FA	-9.388	0.169	-0.00082
	(2.561)	(0.049)	(0.00023)
GM	-7.374	0.141	-0.00071
	(3.616)	(0.068)	(0.00032)
MM	-4.521	0.080	-0.00041
	(2.993)	(0.054)	(0.00024)
OF	-6.656	0.120	-0.00059
	(3.054)	(0.058)	(0.00027)
SC	-2.077@	0.033@	-0.00020@
	(8.062)	(0.150)	(0.00070)
ST	-3.580@	0.068@	-0.00040@
	(2.960)	(0.052)	(0.00023)

Standard Errors in parentheses.
@ indicates coefficients not significant at the .01 level.
All other coefficients are significant at .01 or better.

Table 3.16. Attrition and Training Cost Effects of Selection and Classification Policies

	CURRENT STANDARDS		PLUS 5		PLUS 10	
	ATTRITION	TRAINING COSTS	ATTRITION	TRAINING COSTS	ATTRITION	TRAINING COSTS
RANDOM	791	16.5	617	12.9	368	7.7
CURRENT	0	0	-225	-4.7	-550	-11.5
EPAS	-656	-13.7	-941	-19.7	-1371	-28.7
OPTAACL	-1824	-38.2	-2154	-45.1	-2559	-53.5
OPTAASC	-1824	-38.2	-2253	-47.1	-2834	-59.3
OPTPRFCL	-1733	-36.2	-2020	-42.2	-2422	-50.7
OPTPRFSC	-1733	-36.2	-2123	-44.4	-2651	-55.4
MAXAACON	-1167	-24.4	-1437	-30.1	-1809	-37.8
MAXAAFREE	-2289	-47.9	-2639	-55.2	-3188	-66.7
OPTFLS	-805	-16.8	-907	-19.0	-1181	-24.7
OPTFLSQC	-871	-17.6	-937	-19.5	-1213	-24.9

attrition prediction equations we use here exclude a number of important predictors of attrition (e.g., high school degree status, length of time in DEP). A more complete specification of the attrition prediction equations would undoubtedly increase the potential gains.

Note that, in general, the attrition effects of the OPTFLS alternative are smaller than those produced by the other optimal allocations. This is to be expected since our predictor of attrition is AA score, rather than the FLS predictor of performance. (A regression of the FLS predictor on attrition was done, but the resulting models were significantly less accurate than the equations we use here.) As we shall see in the cost-benefit tables, this effect is more than offset by the value of the improved performance produced by the FLS predictor, but it may be that a policy employing some combination of the FLS predictor and a more complete predictor of attrition than the one we have used here could provide significant gains over any of the alternatives we have simulated.

4. Estimated Recruiting Costs

Table 3.17 summarizes the recruiting cost assumptions and the selection effects of alternative selection scenarios. The average cost function is the same for all scenarios, and

Table 3.17. Average I-IIIA Recruiting Costs Under Current Assignment System Using Three Alternative Cost Assumptions

Selection Standard	Cost Assumption[a]		
	A[b]	B[c]	C[d]
None	5926	5926	5926
Current	8371	8371	8371
Plus5	8511	8595	9487
Plus10	8858	9066	10458

Note: Assumed average cost for low quality recruits is 2290 under all scenarios.

a Cost assumptions differ with respect to (a) the assumed proportion of rejected population in AFQT categories I-IIIA; or (b) the assumed selection method. The average cost function is the same under all scenarios. Increased average costs for the Plus5 and Plus10 scenarios result from the increased number of high quality applicants that must be attracted.

b A assumes that the proportion of I-IIIA in the accession pool remains at 55% under all selection standards, and that increased standards are met by screening applicants on the basis of AA scores. This option assumes that 15% of rejected applicants are I-IIIA, 85% IIIB or IV.

c B also assumes AA-based selection, but that 20% of rejectees are I-IIIA, 80% IIIB or IV.

d. C assumes that current selection practices are used -- that is, the increased standards are met by increasing the proportion of I-IIIA recruits. Under this assumption, the proportion of I-IIIA rises to 63% under the Plus5 option, and to 67% under the Plus10 option.

the average cost of recruiting additional individuals in AFQT category I-IIIA increases as the selection standard rises and more applicants are rejected. The three cost assumptions differ with respect to the assumptions of the rejection rate for applicants in different test categories and how the selection policy is implemented. Cost assumptions A and B both assume that the proportion of accessions that is high quality (I-IIIA) remains constant under all selection standards, and that increased standards are met by increasing the number of applicants processed to find those applicants who are qualified under the increased standard. Under these assumptions, the average recruiting costs change as a function of three variables--the overall rejection rate, the attrition rate, and the proportion of rejectees who are high quality. Assumptions A and B differ only with respect to the assumed proportion of rejectees who are high quality. Assumption A is based on a rate of 15%, and Assumption B uses a rate of 20%. These rates fall on either side of the proportion of total requirements for which high-quality applicants were unqualified under the Plus10 standard in our sample, which was 17%.

Cost assumption D assumes a different selection practice is used. That is, selection standards are increased by recruiting a greater proportion of I-IIIA candidates. Under the Plus5 option the proportion of I-IIIA recruits rises to 63%, and increases to 67% under the Plus10 option. This assumption produces the highest estimates of the costs of increased standards, but also the one most consistent with current practices.

Note that all three assumptions produce the same cost estimates under the current selection standard, and because we assume that "random" selection yields an accession pool that is 50% high quality and no applicants are rejected, the estimates are also the same for the base case of random selection and assignment. The estimates diverge only for the Plus5 and Plus10 scenarios.

Table 3.18a shows how recruiting costs would change under alternative selection strategies under all three cost assumptions. The alternative of no selection standard with random assignment, if implemented, would reduce the number of high quality applicants selected and reduce recruiting costs by 206.3 million dollars. Under current selection standards a random assignment policy would increase attrition, leading to 12.4 million dollars in higher recruiting costs. Lower attrition under EPAS and other assignment strategies would result in lower recruiting costs compared to current selection and assignment procedures.

Table 3.18. Effect of Selection and Classification Policies on Recruiting Requirements and Costs Under Cost Assumptions A, B, and C

SELECT STD	ASSIGNMENT STRATEGY	CHANGE IN HIGH QUALITY APPLICANTS			AVERAGE COST OF HIGH QUALITY			CHANGE IN LOW QUALITY APPLICANTS			CHANGE IN RECRUITING COSTS		
		A	B	C	A	B	C	A	B	C	A	B	C
NONE	**RANDOM**	-10025	-10025	-10025	5926	5926	5926	11626	11626	11626	-206.3	-206.3	-206.3
CURRENT	**RANDOM**	467	467	467	8480	8480	8480	324	324	324	12.4	12.4	12.4
	CURRENT	0	0	0	8371	8371	8371	0	0	0	0.0	0.0	0.0
	EPAS	-387	-387	-387	8280	8280	8280	-269	-269	-269	-10.3	-10.3	-10.3
	OPTAACL	-1076	-1076	-1076	8118	8118	8118	-748	-748	-748	-28.4	-28.4	-28.4
	OPTAASC	-1076	-1076	-1076	8118	8118	8118	-748	-748	-748	-28.4	-28.4	-28.4
	OPTPRFCL	-1022	-1022	-1022	8131	8131	8130	-711	-711	-711	-27.0	-27.0	-27.0
	OPTPRFSC	-1022	-1022	-1022	8131	8131	8130	-711	-711	-711	-27.0	-27.0	-27.0
	MAXAACON	-689	-689	-689	8209	8209	8209	-478	-478	-478	-18.2	-18.2	-18.2
	MAXAAFREE	-1351	-1351	-1351	8053	8053	8053	-938	-938	-938	-35.6	-35.6	-35.6
	OPTFLS	-475	-475	-475	8260	8260	8259	-330	-330	-330	-12.6	-12.6	-12.6
	OPTFLSQG	-514	-514	-514	8250	8250	8250	-357	-357	-357	-13.6	-13.6	-13.6
PLUS5	**RANDOM**	1065	1307	5369	8620	8704	9607	4390	4148	-4752	36.9	42.4	128.4
	CURRENT	597	837	4838	8511	8595	9487	3983	3742	-5063	24.1	29.6	113.5
	EPAS	199	438	4386	8418	8502	9384	3636	3397	-5327	13.3	18.7	100.8
	OPTAACL	-531	-295	3620	8246	8330	9209	3001	2765	-5774	-6.4	-1.0	79.6
	OPTAASC	-476	-239	3557	8259	8343	9195	3049	2813	-5810	-4.9	0.5	77.8
	OPTPRFCL	-401	-164	3704	8277	8361	9228	3114	2877	-5724	-2.9	2.5	81.9
	OPTPRFSC	-458	-222	3639	8264	8347	9213	3064	2828	-5762	-4.4	0.9	80.1
	MAXAACON	-77	161	4072	8353	8437	9312	3396	3158	-5509	5.9	11.3	92.1
	MAXAAFREE	-745	-510	3314	8196	8279	9139	2814	2579	-5953	-12.1	-6.8	71.1
	OPTFLS	*218*	*456*	*4407*	*8422*	*8506*	9389	3652	3413	-5314	13.8	19.2	101.4
	OPTQFLSQG	211	447	4388	8420	8503	9384	3642	3408	-5325	13.6	19.0	100.9
PLUS10	**RANDOM**	2616	3220	9780	8979	9188	10594	9822	9218	-9412	89.1	103.1	239.8
	CURRENT	2093	2692	9166	8858	9066	10458	9335	8736	-9716	74.5	88.3	221.7
	EPAS	1625	2220	8616	8750	8957	10336	8900	8305	-9987	61.5	75.2	205.6
	OPTAACL	791	1379	7821	8556	8762	10158	8125	7537	-10380	38.5	51.9	182.5
	OPTAASC	948	1537	7637	8592	8799	10117	8270	7682	-10471	42.8	56.3	177.2
	OPTPRFCL	1026	1615	7913	8611	8817	10179	8343	7753	-10335	44.9	58.4	185.2
	OPTPRFSC	896	1484	7759	8580	8786	10145	8222	7634	-10410	41.4	54.8	180.7
	MAXAACON	1376	1968	8323	8692	8898	10270	8668	8076	-10132	54.6	68.2	197.1
	MAXAAFREE	589	1175	7400	8509	8714	10064	7937	7351	-10588	33.0	46.3	170.4
	OPTFLS	1734	2330	8743	8775	8982	10364	9001	8405	-9924	64.5	78.2	209.3
	OPTQFLSQG	1693	2289	8722	8765	8973	10359	8978	8369	-9935	63.4	77.1	208.7

Under the Plus5 standard, the direction of the change in recruiting costs depends on both the cost assumption used and the allocation strategy. Under Assumption A, all of the batch optimal assignment policies except those using the FLS predictors, as well as the unconstrained "top-down" policy, reduce attrition by a sufficient amount to offset the higher rejection ratio. Under Assumption B, recruiting costs increase under all strategies except OPTAACL and MAXAAFREE (which produce the highest average AA scores). Under Assumption C, the cost of increasing the I-IIIA proportion of accessions from 59% to 63% causes recruiting costs to increase under all options, with the size of the increase ranging from 71 to 128 million dollars.

The Plus10 option would result in substantial increases in recruiting costs under all allocation strategies. The estimated increases over current costs range from $38 million to $89 million under the lowest cost assumption, and from $170 million to $240 million under Assumption C.

5. Estimated Net Present Value of the Simulated Policies

Table 3.19 provides our estimates of the "gross' value of performance gains under each alternative, as well as the "net" value under each of the three recruiting cost assumptions. The "gross" values shown here are the estimated present values of each alternative prior to accounting for the changes in recruiting and training costs produced by changes in selection ratios and attrition rates. The "net" values are the estimates produced after changes in training and recruiting costs have been accounted for.

The gross value of the performance gains produced by current selection and classification policies is about $325 million dollars annually. However, when the large reduction in recruiting costs that could be realized by moving to a 50% high-quality accession pool are taken into account, the gains provided by the current system drop to just over $150 million annually.

Even under the handicap of predictor score optimization, rather than predicted performance optimization, EPAS provides significant gains over the current system under all three selection standards, with estimated gains under current selection standards of over $56 million annually.

The interaction between efficient allocation and selection standards is apparent in the results under the Plus5 and Plus10 scenarios. Under Cost Assumptions and B, all policies

Table 3.19. Gross and Net Present Value[a] of Change in Expected Performance Under Three Alternative Replacement Cost Assumptions[b]
(All dollar values in millions per year)

SELECT STD	ASSIGNMENT STRATEGY	MEAN PRED PERF	PRED ATTRIT	GROSS VALUE	TRN COST	RECRUIT COST (A)	RECRUIT COST (B)	RECRUIT COST (C)	NET VALUE (A)	NET VALUE (B)	NET VALUE (C)
NONE	RANDOM	0	1601	-325.2	33.5	-206.3	-206.3	-206.3	-152.4	-152.4	-152.4
CURRENT	RANDOM	0.189	791	-21.4	16.5	12.4	12.4	12.4	-50.4	-50.4	-50.4
	CURRENT	0.197	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	EPAS	0.221	-656	32.5	-13.7	-10.3	-10.3	-10.3	56.5	56.5	56.5
	OPTAACL	0.236	-1824	73.5	-38.2	-28.4	-28.4	-28.4	140.1	140.1	140.1
	OPTAASC	0.236	-1824	73.5	-38.2	-28.4	-28.4	-28.4	140.1	140.1	140.1
	OPTPRFCL	0.245	-1733	92.6	-36.2	-27.0	-27.0	-27.0	155.9	155.9	155.9
	OPTPRFSC	0.245	-1733	92.6	-36.2	-27.0	-27.0	-27.0	155.9	155.9	155.9
	MAXAACON	0.229	-1167	49.2	-24.4	-18.2	-18.2	-18.2	91.8	91.8	91.8
	MAXAAFREE	0.254	-2289	104.9	-47.9	-35.6	-35.6	-35.6	188.4	188.4	188.4
	OPTFLS	0.340	-805	232.9	-16.8	-12.6	-12.6	-12.6	262.3	262.3	262.3
	OPTFLSQG	0.334	-871	228.8	-18.2	-13.6	-13.6	-13.6	260.6	260.6	260.6
PLUS5	RANDOM	0.209	61'	12.7	12.9	36.9	42.4	128.4	-37.1	-42.6	-128.6
	CURRENT	0.227	-225	47.6	-4.7	24.1	29.6	113.5	28.2	22.7	-61.2
	EPAS	0.242	-941	67.8	-19.7	13.3	18.7	100.8	74.2	68.7	-13.3
	OPTAACL	0.266	-2154	117.8	-45.1	-6.4	-1.0	79.6	169.2	163.9	83.3
	OPTAASC	0.265	-2253	121.6	-47.1	-4.9	0.5	77.8	173.6	168.3	90.9
	OPTPRFCL	0.264	-2020	136.7	-42.2	-2.9	2.5	81.9	181.8	176.4	97.0
	OPTPRFSC	0.269	-2123	143.2	-44.4	-4.4	0.9	80.1	192.0	186.7	107.5
	MAXAACON	0.250	-1437	84.4	-30.1	5.9	11.3	92.1	108.6	103.2	22.3
	MAXAAFREE	0.279	-2639	145.2	-55.2	-12.1	-6.8	71.1	212.5	207.2	129.3
	OPTFLS	0.386	-907	278.3	-19.0	13.8	19.2	101.4	283.5	278.0	195.9
	OPTFLSQG	0.370	-937	265.3	-19.6	13.6	19.0	100.9	271.3	265.9	184.0
PLUS10	RANDOM	0.236	368	56.9	7.7	89.1	103.1	239.8	-39.9	-53.9	-190.6
	CURRENT	0.254	-550	92.7	-11.5	74.5	88.3	221.7	29.7	15.9	-117.5
	EPAS	0.272	-1371	117.3	-28.7	61.5	75.2	205.6	84.5	70.8	-59.6
	OPTAACL	0.293	-2559	180.9	-53.5	38.5	51.9	182.5	195.9	182.5	51.9
	OPTAASC	0.303	-2834	166.3	-59.3	42.8	56.3	177.2	182.8	169.3	48.4
	OPTPRFCL	0.297	-2422	181.4	-50.7	44.9	58.4	185.2	187.1	173.7	46.9
	OPTPRFSC	0.312	-2651	204.2	-55.4	41.4	54.8	180.7	218.3	204.8	78.9
	MAXAACON	0.276	-1809	127.3	-37.8	54.6	68.2	197.1	110.6	97.0	-31.9
	MAXAAFREE	0.316	-3188	205.3	-66.7	33.0	46.3	170.4	239.0	225.7	101.6
	OPTFLS	0.405	-1181	341.6	-24.7	64.5	78.2	209.3	301.8	288.1	157.0
	OPTFLSQG	0.396	-1213	338.2	-25.4	63.4	77.1	208.7	300.2	286.5	154.9

Notes:
a. All values are relative to the CURRENT allocation, under CURRENT selection standards.

b. "Gross" present value is estimated value of performance gains without accounting for changes in training and recruiting costs. "Net" present value is equal to "Gross" value minus these changes.

except random assignment provide positive net gains over current policies under both Plus5 and Plus10 standards, but the magnitude of the gains (under Assumption B) produced by the current allocation system declines from $22.7 million to $15.9 million when we move from the Plus5 to the Plus10 scenario. Under the same cost assumption, the gains from EPAS *increase* slightly from $68.7 million to $70.8 million as standards are raised. The increases from the batch optimizations are proportionally larger.

If it is assumed that increased standards are met by simply increasing the I-IIIA content of the accessions pool, the advisability of increased standards depends even more critically on the efficiency with which the more expensive supply of recruits is used. Under this cost assumption, the performance gains produced by EPAS are insufficient to offset the increased recruiting costs produced by even a five point increase in cut scores. Furthermore, while the gains remain positive for the batch optimizations, the magnitude of the net benefits under all policies becomes smaller as the selection ratio increases. While this does not imply that increased standards are inefficient, it does suggest that if standards were to be increased, it may be necessary to change the way selection standards are enforced. It is also clear that, as human resources become more expensive, it becomes increasingly important to use those resources efficiently.

The final point to be made with respect to these results is that the potential gains from the use of the FLS predictors are extremely large, yield estimated gains of $260 million under current standards, even when current AFQT quality goals are enforced. Note that this estimated gain is over $100 million higher than the net gains provided by the current system over the alternative of random selection. Again, it is not possible to precisely estimate the proportion of these gains that could be retained under an operational EPAS, but if we use the proportional difference between EPAS and its batch optimal counterpart (OPTAACL) as a rough estimate of the effect of moving from batch to sequential assignment, the expected gains from an EPAS using FLS prediction (and minimizing attrition) could easily exceed $100 million annually.

E. OPPORTUNITY COSTS OF CURRENT CLASSIFICATION POLICIES

1. Rationale

The most serious limitation of the net present value method described in the preceding section is the centrality of the assumption one makes about the dollar value of a standard deviation in performance. While there is persuasive empirical evidence that an

assumption of 40% of salary is a conservative estimate, this "rule-of-thumb" approach is nevertheless often perceived as subjective, and therefore unreliable. This problem is exacerbated when the rule is applied to public sector activities where no clear valuation of output is possible.

An alternative to the NPV approach that, in some circumstances may provide more useful information for the decisionmaker is to focus attention on the *cost* of obtaining a given level of performance using existing procedures instead of attempting to directly measure the net value of the gains achieved under different procedures--that is, to focus on the *opportunity cost* of retaining the existing system.

Figure 3.6 illustrates this approach. The curve LL' is a "budget isoquant" representing the set of output levels that can be obtained from various allocations of a recruit population obtained at some fixed cost B. The curve MM' represents the higher level of output that can be obtained with a more expensive recruit pool costing B^* $(B^*>B)$. The lines OO' and PP' are *production* isoquants representing the levels of performance in each of the two jobs required to produce fixed levels of output, Q and Q^*, $(Q^* > Q)$. This line has a slope of -1, reflecting our assumption that the production function is identical and additive across jobs. The point A is the allocation produced by the current system, and the point A^* is an optimal allocation. Note that this point is on both LL' and PP', indicating that this allocation will produce output level Q^*, while the point A will result in the lower level of output Q occurring along the isoquant OO'. The "gross" value of performance gains used in the NPV method focuses on measuring the difference between the values of Q and Q*. The "opportunity cost" approach instead focuses on the change in the budget isoquant required to move from Q to Q^* if the procedures leading to allocation A are unchanged. Our measure of the opportunity cost of the present system is simply the difference between B and B^*.

2. Methodology

Using this approach in the current context, we ask "What would it cost to achieve the levels of performance produced under each evaluated policy if the mechanism used to achieve those gains was to simply increase the numbers of high quality recruits and assign them using the current system?"

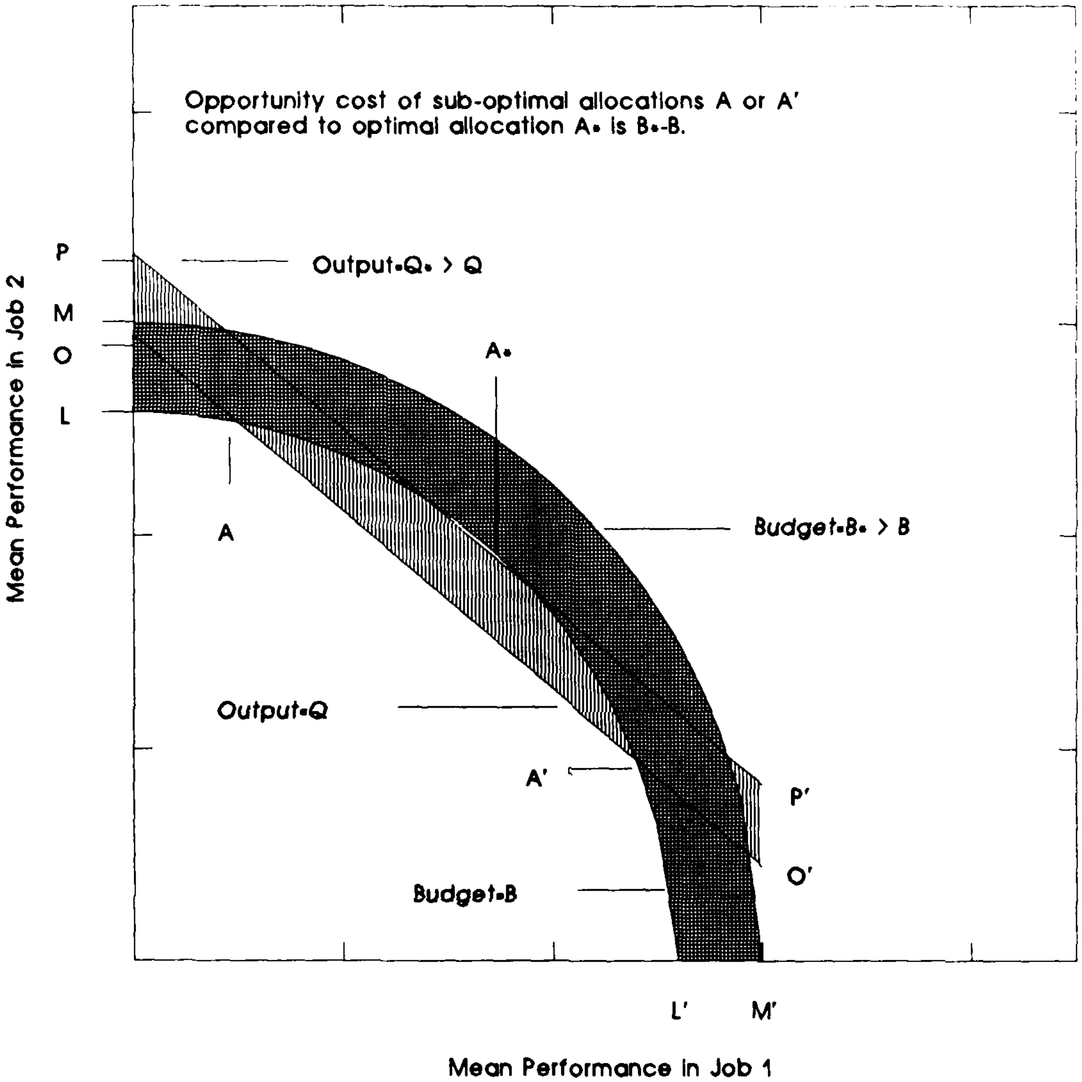


Figure 3.6. Opportunity Cost of Inefficient Allocation

The procedure we used to obtain our opportunity cost estimates was straightforward: We first generated a 50% random sample of 1984 accessions from Army records. (Accessions into MOS not included in our sample were excluded.) We then calculated the FLS prediction of performance in the assigned job for each individual in this sample. Next, we calculated the mean FLS prediction (rounded to two decimal places) for each AFQT percentile level in our sample. Finally, the proportion of I-IIIA recruits required to obtain each of the 23 (rounded) mean levels of predicted performance produced by the simulations (0.19 through 0.42) was obtained by eliminating individuals from the sample, beginning with those with the lowest AFQT scores until the desired mean level of performance among those remaining in the sample was reached. A table containing the resulting sets of I-IIIA percentages and mean performance scores was produced. Finally, the average cost of obtaining each I-IIIA percentage was calculated using the same average cost function used in the NPV analysis, and these costs were added to the table.

Opportunity costs were estimated using the following formula:

OPPCOSTi = {[HQi*ACHi+(1-HQi)*ACL]*(ACC84+DELTATTi)} - COST84,

where OPPCOSTi is the estimated opportunity cost under policy i,

HQi is the required percent of high quality,

ACHi is the associated average cost of high quality recruits,

ACL is the average cost of low quality recruits (again assumed constant at $2290),

ACC84 is the total number of non-prior service accessions in 1984 (120,281), and

DELTATTi is the change from 1984 levels in the expected number of attritions under policy i, and

COST84 is estimated 1984 recruiting costs.

(Note: Interpolation was used to obtain the average I-IIIA cost for mean predicted performance levels falling between the two-digit levels contained in the table.)

3. Results

Table 3.20 shows the results of the opportunity cost analysis. The mean AA scores (in assigned jobs) and mean predicted performance levels are shown in the first two columns of this table. The third column shows the estimated I-IIIA proportion of

Table 3.20. Estimated Cost of Achieving Equivalent Performance by Increasing AFQT CAT I-IIIA Accessions Using Current Selection and Assignment System

SELECT STD	ASSIGN STRATEGY	MEAN AA SCORE	MEAN PRED PERF	REQUIRED PERCENT I-IIIA	AVG I-IIIA COST	CHG IN PRED ATT	CHG IN NUMBER I-IIIA	CHG IN NUMBER IIIB-IV	"OPPOR-TUNITY COST"
NONE	RANDOM	98.3	0	0.46	4649	1601	-14900	16501	-295.6
CURRENT	RANDOM	106.1	0.189	0.58	8142	791	-972	1763	-20.1
	CURRENT	107.5	0.197	0.59	8371	0	0	0	0.0
	EPAS	110.0	0.221	0.63	9195	-656	3559	-4215	81.6
	OPTAACL	113.0	0.236	0.64	9597	-1824	5323	-7147	121.7
	OPTAASC	113.0	0.236	0.64	9597	-1824	5323	-7147	121.7
	OPTPRFCL	112.8	0.245	0.66	9547	-1733	6878	-8611	160.6
	OPTPRFSC	112.8	0.245	0.66	9547	-1733	6878	-8611	160.6
	MAXAACON	111.6	0.229	0.63	9427	-1167	4577	-5744	105.0
	MAXAAFREE	113.9	0.254	0.67	10199	-2289	8002	-10291	187.7
	OPTFLS	108.7	0.340	0.79	13517	-805	23403	-24208	626.1
	OPTFLSQG	109.1	0.330	0.78	13155	-871	21246	-23217	573.0
PLUS5	RANDOM	106.8	0.209	0.61	8910	617	2318	-1701	55.0
	CURRENT	108.7	0.227	0.63	9486	-225	4838	-5063	113.5
	EPAS	110.7	0.242	0.65	9951	-941	6896	-7837	162.8
	OPTAACL	113.8	0.266	0.69	10660	-2154	10080	-12234	241.9
	OPTAASC	114.0	0.265	0.68	10608	-2253	9846	-12099	235.5
	OPTPRFCL	113.5	0.264	0.68	10607	-2020	9840	-11860	235.9
	OPTPRFSC	113.9	0.269	0.69	10774	-2123	10598	-12721	255.6
	MAXAACON	112.3	0.250	0.66	10177	-1437	7905	-9342	187.2
	MAXAAFREE	114.7	0.279	0.70	11058	-2639	11891	-14530	288.9
	OPTFLS	110.2	0.386	0.85	15091	-907	31022	-31929	871.9
	OPTFLSQG	110.3	0.374	0.83	14537	-937	28319	-29256	782.2
PLUS10	RANDOM	107.6	0.236	0.64	9915	368	6735	-6367	161.8
	CURRENT	109.7	0.254	0.67	10458	-550	9166	-9716	221.7
	EPAS	111.9	0.272	0.69	10998	-1371	11617	-12988	284.5
	OPTAACL	114.7	0.293	0.72	11573	-2559	14259	-16818	353.8
	OPTAASC	115.3	0.303	0.74	11886	-2834	15707	-18541	393.7
	OPTPRFCL	114.5	0.297	0.73	11737	-2422	15019	-17441	375.3
	OPTPRFSC	115.2	0.312	0.75	12233	-2651	17326	-19977	410.3
	MAXAACON	113.2	0.276	0.70	11077	-1809	11978	-13787	293.1
	MAXAAFREE	115.9	0.316	0.76	12287	-3188	17580	-20768	446.4
	OPTFLS	110.8	0.405	0.88	15689	-1181	33961	-35142	971.7
	OPTFLSQG	111.1	0.396	0.87	15378	-1213	32430	-33643	918.9

accessions that would be required to achieve the performance level shown in column 2 if the current allocation system was used. The fourth column contains the average cost per high-quality recruit (after accounting for the effect of attrition changes as well as I-IIIA content). The attrition effects are in column 5 and the total required change in the number of I-IIIA is in column 6. The compensating changes in the number of low-quality recruits needed are shown in column 7. The last column shows the change from current levels in the total cost of recruiting this population. (The average costs for this table were computed in the same way as those under the NPV option.) Note that the "opportunity cost" of moving to a mean predicted performance level of 0 are larger than the reduction in recruiting costs under the "random" alternative shown in Table 3.18 ($275 million instead of $206 million). The reason for this is that, if the current allocation system were used, fewer than 50% I-IIIA would be neeaed to achieve the population average level of performance.

In general, the opportunity cost estimates parallel those arrived at under the NPV method in terms of relative magnitudes, but are considerably higher in absolute magnitude. The estimated cost of achieving the performance gains provided by EPAS under current selection standards through the recruitment of additional I-IIIA soldiers is $108.4 million, compared with the NPV estimated gains of $56.5 million. The increased cost of using the current system to achieve the performance provided by the OPTFLS option under current standards would exceed $600 million.

The results for the options involving increased standards should be compared to the CURRENT alternative at each selection level, rather than to the baseline of zero. The deviations from zero are shown only to indicate the estimated costs of increasing standards under the current system. These are roughly $112 million for the Plus5 case and $227 million for the Plus10 case.

The estimated gains of the FLS alternatives relative to the current system increase as standards go up--from $607 million under current standards to $677 million under Plus5 and $741 million under Plus10, which would require an accession pool of 88% I-IIIA.

The relative gains under EPAS decline under the Plus5 standard (from $108 million to $55 million) and then increase slightly (to $64 million) under the Plus10 standard.

F. CONCLUSIONS

The simulation model developed here, together with its accompanying expansion of benefit-cost analysis, provides a number of interesting results.

First, this approach expands the capacity to simulate alternative personnel management policies. Alternative selection, classification, and assignment policies can be simulated in considerably more detail than was possible before. The outcome of these policies can be examined not only against aggregate outcome measures, such as predicted performance and attrition, but can be analyzed in detail by job family or category of recruit. Furthermore, alternative scenarios with different requirements and applicant pools can be readily evaluated.

The approaches to evaluating outcomes has been similarly expanded. We provide two alternative benefit-cost methodologies: one output oriented based upon psychological utility theory, and an alternative input substitution cost approach based on economic substitution theory. Both methods can readily be adapted to new assumptions of training and recruiting costs or SD_y.

The simulations made thus far have produced a number of important findings. First of all, assignment policy can be greatly improved using EPAS. Even when using a different objective from maximizing predicted performance, EPAS produces performance gains worth in excess of $50 million annually. The robustness of these results under alternative benefit evaluation schemes increases our confidence in this conclusion.

The second finding is that it may be desirable to increase enlistment standards. This result must be caveated somewhat, for if recruiting increases in difficulty, or this policy is implemented through simply increasing the proportion of high quality recruits, this policy may not be beneficial. However, if the increased standards can largely be met through screening greater numbers of IIIB and IV applicants, it would be highly beneficial to raise standards.

The third major policy finding of these simulations is that research that improves differential performance is likely to produce substantial net benefits. For example, the full least squares predictor optimization indicates it may be possible to more than double the benefits from assignment. There is likely to be much greater payoffs from psychometric research that improves differential performance than validity research, given the current state of knowledge.

Thus, once a system such as EPAS is implemented, it is likely that there will be substantial payoff from improved classification. The complexities of such procedures as FLS equations would be entirely transparent, since the current operational enlistment and job standards could remain in place.

The results with respect to simultaneous versus sequential selection and assignment are less clear. While simultaneous selection and assignment can provide significant benefits, especially if the selection ratio increases, it appears less important than either improving assignment efficiency or increasing differential prediction.

G. FURTHER RESEARCH

The results from the simulations performed here present several important findings with respect to how selection and assignment can be improved. In addition, the analysis performed here, together with ongoing personnel research, could lead to other improvements. Several of the more prominent possibilities for further research are described here.

The potential for enhancing performance gains through the use of FLS predictors is extremely promising. Further work is needed, however, to obtain more accurate estimates of the performance gains that can be produced by such predictors. For example, it would be useful to perform a full EPAS simulation of such predictors to estimate the gains that could be achieved under more operational conditions that constrained the level of performance gains across occupations.

The simulations performed here rely solely on the ASVAB as a predictor of attrition. Other research, such as Manganaris & Schmitz (1985), has identified other characteristics of recruits that can be used to predict attrition differentially across occupations. For example, education, age, gender, and time in the Delayed Entry Program have all been found to result in attrition differences associated with assignment policy. Nelson and Schmitz (1986) have estimated that substantial additional attrition reductions can be achieved without significantly affecting predicted performance. Thus, it is likely that further assignment benefits can be produced beyond those already simulated.

One area where additional work should be performed is in the area of risk. Decisionmakers need to assess the likelihood that a particular policy would have the desired effect. A policy with a large expected net benefit, but one that incurs substantial risk, may

be undesirable. One of the reasons that we can recommend improved classification, since there are very few risks associated with improving assignment policy.

In order to investigate the risks associated with the kinds of policies investigated here, two things can be done in the future. If a particular policy is being considered, then sensitivity analyses can be performed. All of the key parameters affecting the outcome could be varied until the relative ranking of alternatives changes. For example, since recruiting costs appear to influence the outcome of selection policy, it would be useful to examine at what recruiting cost selection policy changes.

One area that will certainly warrant additional research is the incorporation of new predictors into the personnel system. The Department of Defense is exploring the implementation of new kinds of performance predictors that improve the accuracy with which attrition can be predicted, for example. New tests, particularly ones that are likely to be much less correlated with the ASVAB, would provide considerable opportunity to improve selection and assignment.

Another area that warrants additional investigation is life-cycle manpower modeling. In this chapter we evaluate the benefits of alternatives over one enlistment term. Our approach in theory could be extended beyond one term, perhaps a 20-year military career. It is likely that one may want to evaluate other personnel decisions in separate models, but the same general methodology could be used, or models could be linked to explore cumulative effects of policies over a full career.

One final area that warrants future research is the primary objective of the personnel system. We have explored variations of predicted performance. However, not all kinds of performance gains may be equally valuable. One may find high levels of performance in certain occupations much more desirable. Recent research by Nord and White (1988) has identified such performance value functions for all Army jobs. The impact of including such information in the selection and assignment system should be investigated.

APPENDIX TO CHAPTER 3

Table 3.A.1. Summary Statistics and Predictor Correlations for Simulation Samples: Synthetic Samples with Population Parameters

a. Current Selection Standards

SAMPLE N	3998								
MEAN	105.10	108.59	105.80	106.34	106.69	108.64	108.11	108.14	105.39
STD	11.77	12.09	12.15	11.81	13.21	12.16	11.06	12.68	12.67
CL	--								
CO	0.69	--							
EL	0.89	0.76	--						
FA	0.85	0.86	0.80	--					
GM	0.72	0.78	0.92	0.66	--				
MM	0.52	0.86	0.73	0.65	0.86	--			
OF	0.67	0.88	0.75	0.71	0.84	0.92	--		
SC	0.77	0.91	0.82	0.75	0.87	0.85	0.92	--	
ST	0.86	0.75	0.91	0.79	0.88	0.70	0.82	0.87	--

b. Selection Standards Raised Five Points

SAMPLE N	3943								
MEAN	105.35	108.91	106.09	106.64	107.01	108.96	108.41	108.47	105.71
STD	11.64	11.85	11.97	11.60	13.02	11.92	10.83	12.45	12.45
CL	--								
CO	0.68	--							
EL	0.88	0.75	--						
FA	0.85	0.86	0.80	--					
GM	0.71	0.77	0.92	0.65	--				
MM	0.50	0.85	0.72	0.63	0.86	--			
OF	0.66	0.87	0.73	0.70	0.83	0.92	--		
SC	0.77	0.90	0.82	0.74	0.86	0.84	0.92	--	
ST	0.85	0.74	0.91	0.79	0.87	0.69	0.81	0.86	--

c. Selection Standards Raised Ten Points

SAMPLE N	3780								
MEAN	106.05	109.73	106.93	107.39	107.94	109.79	109.15	109.34	106.57
STD	11.32	11.36	11.46	11.22	12.42	11.43	10.39	11.91	11.94
CL	--								
CO	0.65	--							
EL	0.87	0.72	--						
FA	0.83	0.84	0.78	--					
GM	0.69	0.75	0.91	0.61	--				
MM	0.46	0.83	0.69	0.59	0.84	--			
OF	0.63	0.86	0.71	0.67	0.82	0.91	--		
SC	0.75	0.90	0.80	0.71	0.85	0.83	0.91	--	
ST	0.84	0.72	0.90	0.77	0.86	0.66	0.79	0.85	--

Table 3.A.2. Summary Statistics and Predictor Correlations for Simulation Samples: Synthetic Samples with 1984 Accessions Parameters

a. Current Selection Standards

SAMPLE N	3993								
MEAN	105.62	106.19	106.07	105.92	106.06	106.13	106.28	106.18	106.08
STD	16.76	16.13	16.34	16.36	16.48	16.18	16.00	16.17	16.25
CL	--								
CO	0.71	--							
EL	0.61	0.83	--						
FA	0.77	0.91	0.86	--					
GM	0.53	0.85	0.94	0.76	--				
MM	0.64	0.89	0.82	0.76	0.89	--			
OF	0.76	0.91	0.82	0.82	0.86	0.95	--		
SC	0.94	0.86	0.73	0.81	0.73	0.82	0.90	--	
ST	0.66	0.83	0.94	0.85	0.91	0.81	0.88	0.77	--

b. Selection Standards Raised Five Points

SAMPLE N	3897								
MEAN	106.08	106.73	106.64	106.47	106.62	106.67	106.82	106.70	106.65
STD	16.68	15.93	16.11	16.15	16.26	15.99	15.79	16.01	16.02
CL	--								
CO	0.70	--							
EL	0.60	0.83	--						
FA	0.76	0.90	0.86	--					
GM	0.51	0.84	0.94	0.75	--				
MM	0.62	0.89	0.82	0.75	0.89	--			
OF	0.75	0.90	0.81	0.81	0.86	0.95	--		
SC	0.94	0.86	0.72	0.80	0.72	0.81	0.90	--	
ST	0.65	0.82	0.94	0.84	0.91	0.80	0.87	0.76	--

c. Selection Standards Raised Ten Points

SAMPLE N	3603								
MEAN	107.57	108.36	108.35	108.15	108.26	108.27	108.50	108.29	108.38
STD	16.31	15.38	15.47	15.54	15.75	15.51	15.19	15.51	15.34
CL	--								
CO	0.67	--							
EL	0.56	0.81	--						
FA	0.75	0.89	0.84	--					
GM	0.46	0.83	0.93	0.72	--				
MM	0.59	0.88	0.80	0.73	0.88	--			
OF	0.72	0.89	0.79	0.79	0.84	0.94	--		
SC	0.93	0.84	0.69	0.78	0.69	0.79	0.89	--	
ST	0.61	0.81	0.93	0.83	0.90	0.78	0.86	0.73	--

Table 3.A.3. Average Aptitude Area Scores by Job Family Under Alternative Selection, Classification, and Allocation Policies

SELECTION STANDARD	ASSIGNMENT METHOD	ALL	CL	CO	EL	FA	GM	MM	OF	SC	ST
						JOB FAMILY					
RANDOM	RANDOM	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
CURRENT	RANDOM	106.1	104.5	107.6	105.1	105.4	105.9	107.7	107.3	107.3	104.5
	CURRENT	107.5	100.4	108.7	107.5	104.9	107.3	113.8	105.8	106.4	108.1
	EPAS	110.0	107.9	107.7	112.1	108.1	111.5	111.5	113.5	114.4	110.1
	OPTAACL	113.0	109.7	114.8	112.9	110.4	117.4	113.2	109.2	122.5	113.4
	OPTAASC	113.0	109.7	114.8	112.9	110.4	117.4	113.2	109.2	122.5	113.4
	OPTPRFCL	112.8	110.2	113.3	119.9	108.5	118.7	111.2	112.8	121.9	110.6
	OPTPRFSC	112.8	110.2	113.3	119.9	108.5	118.7	111.2	112.8	121.9	110.6
	MAXAACON	111.6	109.5	113.1	111.5	111.2	118.0	113.7	109.9	111.9	108.5
	MAXAAFREE	113.9	110.1	115.8	113.6	111.7	118.8	113.5	109.1	117.4	115.3
	OPTFLS	108.7	115.6	101.8	113.0	97.8	122.9	109.3	113.1	114.3	105.3
	OPTFLSQG	109.1	107.7	107.5	104.3	94.4	114.5	113.9	113.1	115.0	110.5
PLUS5	RANDOM	106.8	105.1	108.3	105.7	106.1	106.6	108.3	107.9	108.0	105.3
	CURRENT	108.7	102.8	110.3	108.8	106.6	109.2	114.6	107.3	108.3	108.0
	EPAS	110.7	108.5	108.6	112.8	108.9	112.3	112.3	114.5	115.1	110.9
	OPTAACL	113.8	111.4	115.9	114.8	106.1	117.8	114.3	111.2	122.4	113.5
	OPTAASC	114.0	110.5	115.2	113.8	114.5	118.5	115.2	111.1	122.4	113.4
	OPTPRFCL	113.5	110.7	113.3	120.3	112.4	119.1	112.6	113.5	121.2	111.1
	OPTPRFSC	113.9	111.1	113.8	120.8	112.9	119.6	113.1	114.0	121.7	111.5
	MAXAACON	112.3	110.2	113.9	112.1	111.7	118.3	114.5	110.8	111.9	109.1
	MAXAAFREE	114.7	111.0	116.4	114.7	113.5	119.1	114.4	110.7	117.6	115.8
	OPTFLS	110.2	116.1	104.5	113.0	99.3	123.6	110.6	113.2	114.1	108.0
	OPTFLSQG	110.3	109.0	109.1	105.3	96.4	116.1	114.4	113.2	116.4	112.1
PLUS10	RANDOM	107.6	105.9	109.2	106.6	106.8	107.5	109.1	108.7	109.0	106.2
	CURRENT	109.7	104.0	111.2	109.4	107.8	110.5	115.3	108.3	108.8	109.2
	EPAS	111.9	109.3	109.8	114.0	110.1	113.5	113.4	115.7	115.7	111.9
	OPTAACL	114.7	112.1	116.5	115.7	108.8	118.5	115.0	111.8	122.5	114.7
	OPTAASC	115.3	111.9	116.6	115.3	116.9	119.1	116.1	112.7	122.6	114.4
	OPTPRFCL	114.5	112.5	115.9	120.4	103.9	118.1	113.5	114.4	121.1	113.6
	OPTPRFSC	115.2	113.2	116.6	121.1	104.5	118.8	114.2	115.1	121.8	114.3
	MAXAACON	113.2	111.1	114.6	113.2	113.1	118.9	115.5	111.4	111.4	110.0
	MAXAAFREE	115.9	112.5	117.6	116.3	115.3	119.9	115.4	112.2	118.3	116.5
	OPTFLS	110.8	116.4	105.5	113.2	99.6	123.7	111.5	113.3	114.1	108.7
	OPTFLSQG	111.1	109.8	110.1	106.1	97.5	117.0	115.1	113.3	117.6	113.0

Table 3.A.4. Average Predicted Performance by Job Family Under Alternative Selection, Classification, and Allocation Policies

SELECTION STANDARD	ASSIGNMENT METHOD	JOB FAMILY ALL	CL	CO	EL	FA	GM	MM	OF	SC	ST
RANDOM	RANDOM	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
CURRENT	RANDOM	0.189	0.188	0.178	0.241	0.255	0.204	0.209	0.208	0.205	0.133
	CURRENT	0.197	-0.012	0.198	0.274	0.174	0.197	0.364	0.122	0.222	0.229
	EPAS	0.221	0.238	0.127	0.312	0.108	0.244	0.269	0.336	0.353	0.214
	OPTAACL	0.236	0.185	0.258	0.215	0.024	0.320	0.268	0.099	0.455	0.311
	OPTAASC	0.236	0.185	0.258	0.215	0.024	0.320	0.268	0.099	0.455	0.311
	OPTPRFCL	0.245	0.211	0.227	0.453	0.006	0.380	0.211	0.236	0.453	0.234
	OPTPRFSC	0.245	0.211	0.227	0.453	0.006	0.380	0.211	0.236	0.453	0.234
	MAXAACON	0.229	0.204	0.233	0.226	0.099	0.364	0.304	0.162	0.245	0.205
	MAXAAFREE	0.254	0.206	0.264	0.149	0.118	0.390	0.290	0.113	0.412	0.338
	OPTFLS	0.340	0.555	0.037	0.662	0.482	0.722	0.228	0.483	0.463	0.231
	OPTFLSQG	0.330	0.248	0.203	0.319	0.362	0.439	0.371	0.488	0.459	0.375
PLUS5	RANDOM	0.209	0.212	0.198	0.264	0.270	0.226	0.228	0.231	0.228	0.153
	CURRENT	0.227	0.067	0.234	0.288	0.191	0.235	0.386	0.163	0.248	0.226
	EPAS	0.242	0.259	0.148	0.332	0.121	0.265	0.291	0.368	0.370	0.233
	OPTAACL	0.266	0.250	0.291	0.281	-0.049	0.335	0.299	0.170	0.452	0.317
	OPTAASC	0.265	0.210	0.270	0.241	0.118	0.352	0.326	0.162	0.447	0.311
	OPTPRFCL	0.264	0.229	0.227	0.463	0.081	0.389	0.255	0.264	0.428	0.246
	OPTPRFSC	0.269	0.233	0.232	0.472	0.083	0.397	0.260	0.269	0.437	0.251
	MAXAACON	0.250	0.232	0.254	0.248	0.109	0.373	0.328	0.193	0.245	0.222
	MAXAAFREE	0.279	0.238	0.282	0.182	0.156	0.398	0.315	0.167	0.418	0.352
	OPTFLS	0.386	0.575	0.118	0.662	0.540	0.741	0.265	0.492	0.440	0.303
	OPTFLSQG	0.370	0.303	0.250	0.361	0.445	0.485	0.385	0.493	0.498	0.420
PLUS10	RANDOM	0.236	0.242	0.223	0.294	0.292	0.254	0.252	0.261	0.257	0.179
	CURRENT	0.254	0.096	0.259	0.312	0.195	0.277	0.400	0.203	0.276	0.260
	EPAS	0.272	0.284	0.179	0.371	0.147	0.296	0.322	0.411	0.390	0.256
	OPTAACL	0.293	0.273	0.311	0.314	0.010	0.360	0.320	0.193	0.450	0.349
	OPTAASC	0.303	0.261	0.307	0.292	0.177	0.373	0.352	0.219	0.453	0.338
	OPTPRFCL	0.297	0.296	0.296	0.455	-0.053	0.349	0.278	0.291	0.417	0.318
	OPTPRFSC	0.312	0.311	0.312	0.478	-0.056	0.367	0.292	0.306	0.438	0.334
	MAXAACON	0.276	0.264	0.276	0.285	0.139	0.397	0.359	0.214	0.228	0.244
	MAXAAFREE	0.316	0.290	0.316	0.235	0.197	0.422	0.347	0.218	0.440	0.374
	OPTFLS	0.405	0.587	0.147	0.665	0.561	0.746	0.296	0.500	0.440	0.324
	OPTFLSQG	0.396	0.334	0.280	0.387	0.486	0.513	0.406	0.501	0.532	0.445

CHAPTER 4. OPERATIONAL IMPLICATIONS OF THE SIMULATION RESULTS

This chapter addresses issues concerning our simulation, including utility, policy constraints, generalizability of findings, possible changes in policy, and implementing procedures. We start with continued discussion of productivity gains attributable to simultaneous changes in job entry standards and assignment policy using ASVAB.

A. GAINS IN THE CURRENT AND OPTIMAL ASSIGNMENT SYSTEMS

In this section we highlight some of the productivity gains that were described in Chapter 3 and shown in Table 3.19. The productivity gains reported represent dollar-valued gains either over random selection and assignment or over random assignment alone.

The gains represent the values of initial assignment decisions based on prediction of performance and attrition of new recruits for their service during the first tour of duty. Since about the same number of recruits are accessioned into the Army each year, initial assignment decisions result in the same productivity gains each year; consequently, the reported productivity gains are *per year gains* attributable to various assignment strategies using ASVAB.

Although the Army has a well established selection and classification procedure in place based on numerous validity studies, we highlight, in this section, dollars gains over random selection and assignment or over current operational assignment and random assignment. This is done because policymakers frequently pose the question of the utility--dollar value, not validity--of recruitment testing for selection, and because both policy-makers and scientists sometimes question the value of the operational classification system. The more crucial system changes addressed in our simulation are those that appear most promising when psychometric theory and prior results are considered. While all these changes show appreciable gains in utility, they also show considerable variation in their practical feasibility for immediate implementation. The difference in utility among such alternatives is shown in the tables of Chapter 3 and directly reflect the net dollar gains to be

realized by any change being considered. Net dollar differences among alternatives considered are also readily obtainable from the comparisons highlighted in this section.

1. Comparison of Assignment Gains for the Present and Optimal Assignment Strategies

We first consider the gains attributable to the present assignment strategy and the optimal assignment strategy for present job standards and for job standards that are raised by five or ten standard points above the current minimum cut score.

An optimal assignment system provides a realistic upper bound estimate of simultaneous selection and assignment gains that reflect only the constraint of incorporating the requirement of meeting job quotas but not other existing policy and management goals (e.g., quality distribution goals). Consequently, it is not realistic to assume that the optimal assignment strategy would be used operationally; the gains reported here for this strategy, then, represent a gauge that reflects the "cost" of imposing existing requirements and policy constraints on the assignment system.

It is interesting to note that despite these constraints the present assignment system still provides:

- a productivity gain of $50.4 million per year over random assignment
- a productivity gain of $56.5 million by employing EPAS, an efficient computer-based allocation system that meets requirements and policies.

In comparison, an optimal assignment system (i.e., the use of full LSEs) provides a productivity gain of $262.3 million per year over random assignment compared to $50.4 million for the current assignment system, or a gain of 5.2 times more than the gain for the current assignment system.

Before turning to a more feasible and realistic assignment strategy than the optimal system--one that provides upper bound estimates for purposes of comparison--we must consider the effect on the present assignment strategies of raising job standards across jobs by five points taking into account realistic replacement costs--the additional recruiting costs incurred by raising the minimum cutting scores for assignment to jobs. As mentioned in Chapter 3, the results depend on the assumptions made and the allocation method used. In the comparisons below, for example, we used a "medium" replacement cost. (See Table 3.19.)

A five point increase in standards results in:

- a productivity gain of \$22.7 million per year over random assignment for the current system compared to \$278.0 million for the optimal assignment system, or a gain of 12.2 times more than the gain for the current assignment system.

We now consider the effect on the present and optimal assignment strategies of raising job standards across jobs by ten points taking into account replacement costs. A ten point increase in job standards results in:

- a productivity gain of \$15.9 million over random assignment for the current assignment system compared to \$288.1 million for the optimal assignment system, or a gain of 18.1 times more than the gain for the current assignment system.

As noted above, the gains shown here for the five and ten point increases in standards are based on "medium" replacement cost estimates. Proportional gains are obtained for the "low" option A replacement cost estimates as well. However, if we were to use assumptions and estimates made for the "high" option C, the most conservative set of assumptions and estimates made, increasing job standards ceases to be an attractive assumption. We believe that the low and medium options assume more rational and effective recruiting practices than the high option, i.e., that replacements would represent the same "quality" range as used in the original sample, rather than a higher quality range used in making the most conservative estimates. It also assumes that the more effective recruiting practices can be enforced.

Additionally we were unable to evaluate in our simulation an attractive alternative recruiting strategy that would allow recruits to be accepted on the basis of meeting one or more minimum job standard cut scores, even though they normally would have been rejected on the basis of AFQT scores under current practices. In a later chapter we describe a multidimension screening (MDS) system that calls for selection and classification decisions to be made simultaneously. Minimum cut scores are used only as "basement" standards. If the concepts of simultaneous decisions were to be used on an interim basis before implementing MDS, utilizing individuals that meet raised standards on one or more assignment composites should result in productivity gains for current practices. It requires a model sampling experiment to confirm this expectation.

2. Comparison of Selection and Classification Gains for the Present and Optimal Assignment Strategies

The last consideration to be highlighted in this section is the simultaneous effect on selection and classification (assignment) of the current and optimal assignment strategies

using present and raised jobs. It is interesting to note that the *gross* value of productivity gains for selection is $325.2 million; when we account for changes in training and recruiting costs the *net* value for selection is $152.4 million. The large difference between gross and net is attributable to the high, but realistic cost of recruitment using the current selection ratio. (See Table 3.19.) Other results show:

- the productivity gain of the current system is $202.8 million per year over random selection and assignment using present job standards for assignment compared to $414.7 million for the optimal system
- the productivity gain of the current system is $175.1 million over random selection and assignment using a five point increase in job standards compared to $430.4 million for the optimal system.

When low estimates of replacement costs are used (see Table 3.19, Option A) in place of medium costs, results show relative gains similar to those for medium cost. Gains are always greatest for the ten point increase for the optimal system, next for the five point increase, an smallest for the current system; the optimal system gain always exceeds the current system gains for each job standard evaluated. As noted in the section above, the C option provides a different pattern of results.

B. GAINS RESULTS FROM A MODIFIED OPTIMAL ASSIGNMENT SYSTEM

Again, the optimal assignment system provides our upper bound estimate of selection and assignment productivity gains if the only external requirement imposed on the system is to meet job quotas (i.e., filling job slots in each MOS with the required number of enlistees). If all operational requirements and policy and management goals are to be met, the optimal assignment strategy would need to be "constrained" to satisfy these demands explicitly, as does the current assignment system.

Table 3.12 shows an average gain in predicted performance of 0.197 standard deviation units for the current assignment system over random assignment, compared to a gain of 0.340 for the optimal assignment system, or a gain of 73 percent over the current system. The reduction in the optimal assignment system that results by meeting most major constraints or requirements including quality goals is about 2 percent.

This new strategy, the "constrained LSEs assignment system" system, called "OPTFLSAQG" in Table 3.19, shows a gain of 0.334 standard deviation units or a gain of about 70 percent over the current system using present job standards.

1. Comparison of Assignment Gains for the Present and Constrained LSEs Strategies

The productivity gains for the constrained LSEs assignment strategy are about 2 percent less than for the optimal strategy. The gains for medium replacement costs show:

- the productivity gain for the current assignment system remains the same at $50.4 million per year over random assignment compared to $260.6 million for the constrained LSEs assignment system, or a gain of 5.1 times more than the gain for the current system.
- with a five point increase in job standards, a productivity gain $22.7 million per year over random assignment for the current system compared to $265.9 million for the constrained LSEs system, or a gain of 11.7 times greater than the gain for the current system.
- with a ten point increase in job standards, a productivity gain of $15.9 million per year over random assignment for the current system compared to $ 286.5 million for the constrained LSEs system, or a gain of 18.0 times greater than the gain for the current system.

2. Comparison of Selection and Classification Gains for the Present and Constrained LSEs Strategies

We finally consider the simultaneous effect on selection and assignment of the current and constrained LSEs strategies using present and raised job standards across jobs. As expected, the results show a slight reduction in gains for the constrained LSEs:

- the productivity gain for the current assignment system is $152.4 million per year over random selection and assignment using present standards compared to $413.0 million for the constrained LSEs strategy.
- the productivity gain for the current assignment system is $175.1 million per year over random assignment and assignment using a five point increase in job standards compared to $418.3 million for the LSEs strategy.

Again, when low estimates of replacement costs are used (see Table 3.19, Option A) in place of medium replacements costs, results show similar relative gains.

Thus, in examining all the differences between the current system and an optimal system or a constrained optimal system (i.e., constrained LSEs that meet all major requirements and policies), very sizeable gains are achieved by using an optimal system or, more appropriately, a constrained LSEs system. Assignment by LSEs produces gains that are between 5 times and 18 times greater than the gains attained by the current system.

It is these sizeable gains that could be captured by revising the assignment policy now in use. The revisions would only call for better use of the information contained within the present ASVAB and a simultaneous increase in job standards of 5 or more points in minimum cut scores.

3. Extending Results to Other Services

Although the present analysis was confined to the Army's selection and classification system, we believe that the results will generalize to all the military services. It is widely recognized that ASVAB validities and assignment policies and procedures are quite comparable across the services.

If the productivity gains found in the present analysis were extended beyond the Army's accession of 41 percent of recruits to all military services attributable to an optimal selection and classification system, the productivity gain would be about $1.011 billion per year over random selection and classification; for a constrained LSE system, the gain would be about $1.007 billion per year. In contrast, the gain attributable to the current system is $494.6 million.

We believe the above estimates are conservative compared to actual productivity gains because, as discussed later in this chapter, most of the parameter values used in our study were underestimates. For example, our dollar value estimate is based on the very conservative proportional rule--an SD_y estimate equal to 40 percent of salary. Also our estimates appear to be very conservative in contrast to opportunity costs. For example, the Army's use of an efficient selection and assignment system (FLS), under current standards, would result in a productivity gain of $414.7 million compared to the opportunity cost of $640.9 million to just recruit equivalent levels of performance.

It is important to note that there are, of course, further gains possible beyond the first tour of duty for the cohort group depending on length of service or tenure. Furthermore, there are additional gains that are realized beyond tenure because most recruits are promoted and assigned to more complex jobs that have larger associated SD_y values. Thus for the same number of productive man-months of service, productivity gains for the second tour of duty are considerably greater than gains for the first tour of duty although not reflected in the simulation results.

Although we believe that our estimates, taken as a whole, provide a conservative estimate of utility gains attributable to selection and classification, we note again that two

estimates of assignment gains were more likely to be overestimates. As described in Chapter 3, the gain attributed to the optimal assignment strategy included some correlated sampling error because the weights given to the assignment variables were the same as those given to the evaluation variables. However, because our "sample" used to computer weights for both assignment and the evaluation variables was based on the combination of several very large samples, and the variation in validities reduced by moving outlines toward the mean, we believe the effect on the measurement of mean predicted performance was negligible.

C. POLICY AND PROCEDURAL CONSTRAINTS IN THE SIMULATION

The Army's job assignment systems, as well as those of the other services, are constrained by an extensive set of policy and managerial considerations detailed in Chapter 2. The more confining these constraints become, the smaller become the differences among feasible alternative assignment systems as measured by predicted performance or utility. Nord and White (1988) summarize these constraints and their implications:

> [Constraints] include not only limitations imposed by force structure requirements and the availability of training resources, but also a number of policy constraints whose purpose is to insure an acceptable, if not optimal distribution of performance across jobs. This latter set of constraints includes minimum job entry standards, an MOS priority system, and a set of job-specific "quality goals" based on educational attainment and scores on the Armed Forces Qualification Test (AFQT).... One of the effects of these constraints, when they are used in optimal assignment, is to mitigate the effects of variation in validity and job quotas--producing an allocation in which average performance is lower, but also less variable across jobs than would occur without them.
>
> If one assumes that these requirements have evolved in order to enhance Army productivity, then their existence implies two things: (a) that job performance is not equally valuable at all levels in all jobs; and (b) that the payoffs to increases in performance tend to decline in most jobs as the average level of performance increases.
>
> The first conclusion is implied by entry standards, the variation in which is based on the fact that low levels of performance are more tolerable in some jobs than in others. The second is implied by the existence of quality jobs, which have two effects: First, differences in goals across MOS imply job-specific differences in the value of high level performance. Second, the role of the goals as constraints in the assignment process has the effect of reducing the payoffs to high-performance assignments in jobs where quality goals are approximately satisfied and increasing these payoffs in jobs that are falling short of the goals. (p. 10).

Thus, in generating alternatives for consideration in a modified assignment system it is important to consider their policy implications and the readiness of policymakers to accept changes called for by use of such alternatives.

1. The Use of FLS Composites Without Hierarchical Effects

The productivity gains highlighted above all require some modification of existing policy. One alternative policy that would not require a policy change would merely substitute FLS composites in Army Standard score form, i.e., with hierarchical effects removed, in place of the existing AA composites. Unfortunately this effective, practical and feasible alternative, the closest alternative to the current operational system, was not considered in time to be specifically addressed in our simulation.

It may be recalled from Chapter 1 that classification efficiency has two sources, allocation efficiency and hierarchical classification efficiency. The allocation process capitalizes on differential validity (broadly defined). All classification efficiency not explainable as hierarchical classification--that resulting from disparate means and variances of criterion variables across jobs--is attributable to allocation efficiency. When heterogeneous validities and/or job values (importance or criticality) are attached to jobs and are also reflected in the predictor variables used in the assignment process, hierarchical layering effects result. This hierarchical layering can provide substantial hierarchical classification efficiency.

Least squares regression weights (LSEs) applied to all tests of the ASVAB forming test composites corresponding to each job family and a general composite that predicts performance in all jobs provide maximum utility when used in both or either selection and classification. Such composites will not only provide the means of maximizing average validities across jobs, but will also maximize potential allocation efficiency (PAE). The validities of the job family specific composites are multiple correlation coefficients between the composites and each job criterion measure. The validity of the "general" composite is the multiple correlation coefficient computed using the weights that are best when the analysis sample is the aggregation of all job samples. All of the tests that are used for the nine job family-specific LSEs are also the best LSEs selection instruments for use in separate, direct selection of applicants for each job family. If the composites use a reduced number of tests or are not LSEs, the best composites for selection are not necessarily the best for classification. In brief, the LSEs maximize both predictive validity (PSE) and the

PAE obtainable from the battery whenever the LSEs are based on all tests in the operational battery--and an optimal selection/assignment algorithm is used.

A difference among mean benefit scores across jobs may result entirely from differences among the predicted performance scores; these differences result from either differences in validities or from the differences in value accorded to jobs (both differences may exist in the same situation). To capitalize on differences in validities (i.e., hierarchical classification), the most effective composites to be used as assignment variables are the least squares predictor(s) of the predicted benefits since this is also classification, albeit hierarchical classification.

The current Army aptitude area composite predictors, using an optimal assignment algorithm, does not in any way capitalize on the hierarchical laying effect since the composites were standardized to have equal means and variances and are not weighted by either validity or job values. Therefore the Army's current use of aptitude area composites as assignment variables relies only on PAE (with no hierarchical classification effects); since LSEs are not used they do not maximize predictive validities nor PAE, and for the reasons given above have no hierarchical classification efficiency.

The LSE composites using every test in the battery as an independent variable are referred to as full least squares (FLS) composites. Such FLS composites, standardized and divided by their multiple validity coefficients to obtain equality of means and variances, will provide allocation efficiency without any capability of capitalizing on hierarchical layering. We estimate that such FLS composites converted to Army standards scores, i.e., removing the effects of varying validities, would still result in about half the productivity gains achieved through the use of predicted performance (PP) measures, the optimal FLS composites, that capitalize on both hierarchical layering and allocation efficiency.

Again the advantage is considering FLS composites converted to Army standard scores is that the existing AA composites could be replaced without changing policy or any existing procedure. The complexities of such FLS composites would be entirely transparent to operational personnel once a score has been computed and provided the traditional AA label (CL., etc.) and would become visible to the applicant-recruit, recruiter, counselor, or personnel clerks, since the current operational system would remain in place.

2. Substitute for AFQT-based Quality Goals

A second policy alternative that is worthy of serious consideration and that was also not included in our simulation is the evaluation of LSE composites used to achieve "quality goals" in place of AFQT scores. A comparison of alternative strategies under both quality goal conditions would permit the measurement of their relative impact on classification efficiency.

As noted in Chapter 2 the current assignment system includes a set of AFQT-based quality goals. These goals are defined in terms of minimum percentage of AFQT category 1-IIIA accessions in each MOS. In practice, the AFQT is not only the instrument used for selection but the principal assignment instrument as well. Data show that the average test category 1-IIIA recruit qualifies for 96 percent of all Army jobs. Reliance on AFQT for selection and to ensure quality distribution goals for jobs to assign to high quality recruits presently accessioned into the Army reduces the room for obtaining PCE using the current operational aptitude area composites.

While the use of AFQT quality goals contributes to a more balanced distribution of performance levels across jobs and helps ensure that each MOS has a pool of potentially promotable enlistees of high "quality", the AFQT is not the best measure to use for these purposes. AFQT is used, in the context of quality goals, as a measure of general mental ability, but the measure that is most desirable is one that defines "quality" in terms of predicted performance in a job. Our utility study clearly indicates that LSEs are appropriate for this purpose. If a measure of *performance* (LSEs) were to be used in defining quality goals rather than a measure of *ability* we would be able to define the term "quality" consistently and precisely in terms of performance and at the same time predict performance more accurately, particularly for the combat arms where quality goals are relied on more extensively than in other job families.

Some may believe that a measure of general mental ability is needed to "grow" future leaders for later tours of duty. LSEs, based on the present ASVAB, are comprised entirely of cognitive ability measures. Project A, however, has demonstrated that the validity of non-cognitive measures also would add to the predictability of first tour job performance and in all likelihood performance beyond the first tour. The composite that best captures future "leader" performance also will undoubtedly be an LSE composite, not AFQT; the most appropriate measure of "quality" is always mean predicted performance, not general mental ability. Thus LSEs should be used in place of AFQT to define minimum quality distribution goals. The appropriate cutting score for LSEs should be consistent

with quality goals and also should be based on expected or predicted performance, and, if desired, weighted by payoff or value of performance level/job combinations.

As noted in an earlier report (Zeidner and Johnson, 1989) there is an additional advantage in using job family-specific composites for specifying "quality" goals rather than a simple composite. In a sample of 7,500 applicants, 56 percent reached or exceeded the 50th percentile on AFQT, compared to 40 percent who reached or exceeded the average standard score on their best AA. Thus a very large majority of enlistees assigned to jobs on the basis of AA would achieve predicted performance scores as high as or higher than the average score of the entire sample (Maier and Fuchs, 1972). This apparent impossibility in which nearly everyone could be above average is attained by capitalizing on intra-individual differences, for nearly everyone excels in some aptitude (Anastasi, 1988). Using nine LSE composites (one for each job family), rather than one general mental ability composite (AFQT) would result in reducing the need to impose quality goal constraints. This procedure in turn would result in an operational LSEs assignment system closer to an optimal assignment system.

In brief, a higher level of PCE obtainable from the present ASVAB or any future ASVAB should result from the use of LSEs in defining quality distribution goals and in making actual assignments than using an AFQT score, an unweighted, ASVAB composite of general mental ability. A simulation would indicate the extent of increase in PCE in using LSEs over AFQTs scores as quality goal standards, we believe other samples and policies may produce greater gains through the use of LSEs goals.

The goal of a selection and assignment system is to maximize the productivity of the human resource available to the Army. It is evident from historical manpower utilization practices that policymakers believe assignment decisions should not be driven solely by individual performance, but rather by the perceived value or utility of performance. Traditionally, the technical branches and combat arms components compete to procure their fair share of available "quality" enlistees. The technical branches are given only the essential number of quality enlistees proven essential to complete technical training successfully so as to make available as many high quality enlistees as possible to the combat arms. Quality distribution goals and prioritization of job quotas assist in meeting these values.

Because AFQT is used to determine enlistment standards and scores are categorized into "mental" groups, policymakers have become accustomed to think of "quality" in terms

of mental groups as measured by AFQT, a composite of four subtests of the ASVAB. To place a value on jobs or performance level/job combinations is clearly a policy decision. The best way to implement these decisions, both in meeting policy goals and making assignments, is through the use of LSEs.

In the event that policymakers are willing to value performance explicitly, this information can be readily incorporated into the assignment procedure, serving to increase the PCE of the assignment strategy further by utilizing hierarchical layering effects.

The current practice of using quality distribution goals and priority lists serves as a constraint on the potential classification efficiency of the ASVAB. Even with full acceptance of these policies as given, their limiting effects on the assignment process could still be greatly ameliorated. The key to better utilization of personnel is improved assignment procedures through consistent employment of the same underlying concept of quality, the common thread running through selection, quality goals and assignment--mean predicted performance.

3. Improved Prediction of Attrition

As noted in Chapter III, attrition rate affects recruiting, training and salary costs. Our analysis shows that even small changes in the attrition rate are important because of its associated large costs. Previous research (Nelson, 1985; Nelson and Schmitz, 1986) estimated that it may be possible to obtain significant reductions while retaining nearly all of the gain in predicted performance when attrition is added to the objective function.

While our simulation accounted for changes in training and recruiting costs for the new accessions needed to obtain a fixed quantity of "effective man-months," the simulation did not attempt to minimize attrition costs. A simulation which considers both maximizing predicted performance and minimizing attrition as the combined objective function would likely provide further assignment benefits from those strategies already simulated. Research also indicates that ASVAB measures could be augmented by other valid, readily available measures such as age, gender and time in the Delayed Entry Program (Manganaris and Schmitz, 1985).

D. LIMITATIONS IN THE INTERPRETABILITY AND GENERALIZABILITY OF FINDINGS

1. Sample Characteristics

The distribution of characteristics in the empirical sample of 1984 used in our simulation is partially determined by such transient factors as economic conditions that may limit the generalizability of findings based exclusively on this single empirical sample. In order to evaluate the robustness or results, we generated a synthetic population with the same means, standard deviations and expected intercorrelations observed in the empirical sample and also a synthetic sample that was equivalent to the youth population.

The relative magnitudes of MPP across both selection standards and assignment policies were found to be very consistent between the synthetic score samples and the empirical sample, suggesting that the predictions produced from the empirical sample are likely to hold up, in relative terms, under a reasonably wide variation in accession populations. Further, the precise amount of dollar savings is not as important as are the differences in mean predicted performance among alternative strategies. We know from an examination of our tabled results that improvements of one or two tenths of a standard deviation of mean predicted performance may result in very large dollar gains. For example, the constrained optimal LSEs alternative produce a 0.143 gain in mean predicted performance. This gain results in $260.6 million more for the constrained LSEs strategy than for the current assignment composite.

The substantial gain of 0.1433 in MPP of the constrained optimal LSEs is attributable principally to improved allocation efficiency through increased differential validities of prediction composites and to improved hierarchical classification through weighting of composites by their validities.

Contrary to the erroneous belief of many, the average increase of .05 in the multiple correlation over the average single test validity can have made only a minor contribution to the improvement of classification efficiency if not accompanied by other indications of increased PCE.

2. Stability of Least Squares Weights

While an FLS assignment system will always be superior to other types of composites on theoretical grounds, there may be some concern when such composites are used operationally in independent samples. The important concern is not to obtain reliable

estimates of the true weights in each composite; but to obtain stable predictions of performance. As noted in Chapter 3, we examined the stability of the predictors by creating several sets of weights using small perturbations of the intercorrelation matrices. Each set of weights was used to produce a prediction of performance.

The resulting correlations among the different predicted values ranged between 0.95 and 0.99, thus providing some confidence that the predicted performance estimates based on FLS composites are reasonably stable.

When the validities and intercorrelations are computed on extremely large samples and regressed towards the mean, as ours are, it becomes credible that the clear superiority of LSEs for predicting performance in independent empirical samples of the same youth population is a finding that can be expected to be repeated in a more completely controlled design. Note that the simulation sample is entirely independent of the data aggregated and corrected to obtain regression weight estimates that are assumed to represent the youth population. Thus, the traditional shrinkage formulae are not appropriate.

3. Representativeness of Jobs Sampled

A limitation of our data for computing regression weight relates to the number of jobs in our sample and the extent that the jobs are representative of the 260 entry level Army MSO. The job validities we employed for determining mean validity vectors for each job family were based on extraordinarily large samples in: 23 MOS (Maier and Grafton, 1981), 68 MOS (McLaughlin et al., 1984), and 19 MOS (McHenry, 1987; Eaton, 1987). The MOS in these studies represent jobs with large numbers of accessions and/or were considered important or critical by policymakers and were proportionally weighted by the number of operational accessions in each job to enhance representativeness of the sample of jobs.

Although the validities used in our study were based on large samples and carefully developed performance measures (especially those developed for Project A), it would have been desirable to have included validities of more jobs in our study. Nevertheless, because of the magnitude of the cumulated data on hand, we feel that the mean validities of predictors and resulting productivity gains within job families obtained from the simulation sample would not have varied greatly by including more MOS to obtain "universe" regression weights. Additionally and more importantly, from the point of view of decisionmaking, the choice of the LSEs as the best alternative clearly would have remained the same if more jobs within each job family had been represented.

A problem associated with the "representativeness" of jobs within our data relates to research expressly designed to improve the PCE of batteries and composites. Although jobs were generally selected to be as representative of a job family as practical, representativeness accomplished by the inclusion of more densely populated jobs was done at the expense of adequately exploring the full dimensionality of the joint prediction-criterion space. As the number of separate sub-families included in a study becomes more limited (e.g., the 19 MOS in Project A), it becomes more difficult to find the PAE attainable from a battery.

The Army Research Institute has been exploring a synthetic validation approach as a means of designating predictor composites for all MOS by extending the findings on the 19 MOS empirically validated in Project A. (Wise, McHenry, Campbell, and Arabian, 1987). Synthetic validities are needed because of the expense, time and effort required to empirically determine validities for the large number of entry level MOS. Additionally, as new weapons systems are developed that impose different types of job demands on enlistees, new or revised MOS regularly are added and change the job family structure. Also, estimates of validities for these new jobs become very important during the earlier phase of development and design of new systems.

The basic concept of synthetic validity involves the identification of common components for a variety of jobs, the determination of validity of each predictor for each job component performance, and then, after the components of a "new" job are identified, forming the prediction equation for the new job combining valid predictors of each job component. Trattner (1982) delineated four different synthetic validity approaches taken by Guion (1965), Lawshe (1952), McCormick, Jeanneret, and Mecham (1972), and Primoff (1975). More recently, Mossholder and Arvey's (1984) article provided a conceptual and comparative review.

Wise et al. (1987) stated that the success of their effort depends on identifying a set of components that adequately cover important attributes for al enlisted jobs, reasonably predicting performance in these components using ASVAB subtests or experimental measures and combining the separate prediction equations for each component into an overall prediction equation. The authors noted that Project A produced reliable differential prediction of technical proficiency in different MOS (Wise, Campbell and Peterson, 1987). Therefore, they reasoned, it may be possible to group jobs into a number of families based on similarities and differences in prediction equations in the joint predictor-criterion space within and across job family.

In establishing job families, several considerations are worth emphasizing: (1) the joint predictor-criterion space is the only relevant domain for clustering jobs for use in the assignment process; (2) the merging of jobs into families always reduces PCE compared to the use of different LSEs for each job; (3) increasing the number of job families and their corresponding composites increases PCE until the number of families equals the number of jobs; (4) a different job family structure could result from agglutinating jobs to maximize PSE (producing high correlations among LSEs in each cluster) than from agglutinating jobs to maximize PCE (minimizing the differences among LSEs in a cluster in order to maximize LSEs' differences across job clusters); and (5) the reliability of clustering jobs in cross-sample comparisons will not be high unless appropriate weighting is given to core jobs in a job family relative to jobs close to the boundaries of other families.

The objective of clustering jobs into job families with corresponding test composites for use in classification, we suggest, is to maximize differential validity (i.e., H_d or PDI). Johnson and Zeidner (1989) describe job-clustering procedures that elaborate on the considerations listed above to maximize differential validity.

The major need for clustering jobs into families in the assignment process could be removed by the use of FLS composites as the assignment variables for each job instead of for job families.[4] An additional but important need for clustering is for using test composites (possibly but not necessarily with a reduced number of tests) as a practical mechanism in the recruiting and assignment process. While full LSEs for each job may be used to make actual computer-based decisions, a smaller number of LSE composites estimating classification efficient factors that span the joint predictor-criterion space may be useful to recruiters, counselors and applicants as in a two-tiered system discussed further in Chapter 5. Using such composites may aid in understanding and communicating assignment choices and job standards; also, composites may be recorded for use in making future career related decisions, whereas the job-specific FLS composite scores would not be retained in personnel records. The results of ARI's synthetic validation study, when available, warrants a research effort to identify and evaluate a use of optimal composites for these purposes. Such a research effort is described in the next chapter.

---

4 The use of LSEs for job families to adjust the LSEs for jobs having small validation samples would still require the identification of job clusters, possibly a separate cluster centering on each job.

4. The Use of the Forty-Percent Proportional Rule

A limitation in our study is the use of an SD_y estimate equal to 40 percent of salary. This value is widely recognized to be a lower bound estimate which invariably underestimates productivity gains. Estimates for use in our study could have been empirically determined using one or another of the SD_y estimating techniques. In the context of our study, a different SD_y estimate would likely result in higher productivity gains (Eaton et al., 1985); however, the decision as to which is the best alternative would remain the same.

It is of interest to note that Schmidt, Hunter and Dunn (1987) estimated SD_y percentage of 106 percent for high complexity Navy jobs, 63 percent for medium complexity jobs, and 43 percent for low complexity jobs. They placed 14 percent, 70 percent, and 16 percent of Navy jobs at these complexity levels, respectively.

Additionally, as noted in Chapter 3, the results of using opportunity costs as a measure of the benefits produce results that are generally comparable to the 40 percent rule-of-thumb. The relative order and magnitude of the benefits of alternative strategies is also essentially the same, except for the opportunity cost of FLS assignment.

5. The Use of the Technical Proficiency Job Component

Another limitation in the data resulted from the use of only a technical proficiency criterion component. In a particular set of Project A data collected at a later date, job performance as measured is comprised of five very distinct components. The other four job components are basic soldiering skills, leadership and effort, personal discipline, and military bearing and personal fitness. These components may be perceived as proficiency-based and motivationally-based aspects of performance (Wise, Campbell, McHenry and Hanser, 1986). The reliabilities for technical proficiency and basic soldiering were both 0.85; reliabilities for the other three components were all 0.80 (Zeidner, 1987).

Since an overall criterion composite combining all five components was not available on the data set used for our study, we in effect were using only the validities for the technical proficiency component. The use of an overall weighted multidimensional composite may have resulted in a small change in mean predicted performance if the new Project A predictors had also been included. However, Wise, Campbell, and Peterson (1987) showed that the hypothesis that a single best weighted composite fits all jobs could not be rejected separately using each of the five criterion composite measures other than for

technical proficiency. Thus the presence or absence of these "other" four components has more effect on selection than on classification. The use of a criterion that includes only the technical proficiency component appears to increase potential allocation efficiency by a fairly small amount, if at all.

6. The Use of Aptitude Area Validities

Another limitation in our data is that the LSEs employing predictor weights on AA composite scores, as used in this analysis, provide an approximation of the utility obtainable from the use of LSE composites that use weights based on the full ten tests of the ASVAB. Our LSEs weights are based on the validities and intercorrelations of the nine aptitude area (AA) composites rather than on the validities and intercorrelations of the ten tests. Since scores for each subtest were unavailable in the data set utilized in the simulation, using independent variables based on composites with overlapping tests reduces the level of PCE obtainable from the ASVAB. This overlapping of tests is more serious because these AA composites were constituted to maximize predicted validity rather than PCE. Thus again the utility of the FLS composite alternative is underestimated.

7. Valuing Jobs Equally

Another limitation in our data is that the full potential of hierarchical layering effects was not achieved since jobs were equally valued. If jobs were assessed by importance or by trade-offs among different performance level/job value combinations and such information were embedded in both assignment and evaluation variables, mean predicted performance of the FLS alternative would increase.

Nord and White (1988) employed a simulation technique to evaluate the effects on assignment of two alternative strategies, maximizing performance regardless of job performance values and maximizing the utility of performance. Field-grade officers made judgments on payoffs of performance level/job combinations for entry-level MOS during tour of duty. (Sadacca, White, Campbell, DiFazio, and Schultz, 1988). Assignments to maximize performance were found to yield large gains in productivity compared to the current Army assignment system, but produced performance distributions that were highly variable across jobs and sensitive to the interaction between validity and job size. Assignments to maximize utility of performance produced comparable gains to the performance strategy but provided a more balanced distribution of performance across jobs. The authors concluded that to maximize productivity the value of performance should

be incorporated into the assignment system and that such utility trade-offs would improve the precision in evaluating manpower policy alternatives. However, this valuation of jobs would require a change in the traditional policy of valuing jobs equally.

8. Risk Assessment

Another limitation in our data is that we did not subject the parameters used in the analysis to a risk assessment (except in the case of recruiting costs) through the use of Monte Carlo-type techniques such as sensitivity analysis or simulations with perturbations of estimates. While such a risk analysis would most likely result in the selection of the same LSEs assignment strategy (because of the large MPP gain), it would provide a better understanding of how utility values change as a function of parameter variability.

In the case of raising job standards minimum cutting scores by five or ten points, we found that such increases produced gains in the net value of job performance for two of the three estimates of recruiting costs.

Our simulation did not consider the interaction of feasible recruiting strategies and recruit behavior. Since there is a significant relationship between job preference and aptitude, it would have been desirable to simulate, through a model sampling experiment, a number of realistic recruiting strategies and different types of recruit behavior. Once the issue of recruiting strategy was resolved, a simulation could then vary estimates of recruiting costs in a sensitivity analysis to determine a more precise estimate of the optimal minimum job standard cut score.

9. Correlated Error Component

A final shortcoming of the data also addressed in Chapter 3 may result from correlated sampling error due to the use of the same weights for the identical set of assignment and evaluation variable scores in measuring mean predicted performance. Parameters computed on the basis of the data currently available permit us to define the universe of the past decade with some confidence, but with less confidence when estimating the future universe. While we used very large samples in the determination of estimates, it was not feasible in our simulation to measure the effects of correlated sampling error present in both assignment and evaluation variables. The effect is believed to be more than balanced by the underestimates of utility provided by other factors since alternative estimates of the intercorrelations that provided very different looking regression weights

when combined with the validities still provided scores that were very highly correlated across estimates.

10. Combined Effects

Taken together, these shortcomings most likely resulted in a considerable underestimate of productivity gains attributable to the use of a full LSEs assignment policy. A series of model sampling studies are now under way that will remove correlated sampling error (between assignment and evaluation variables) and directly estimate predicted performance from the use of all ten tests of the ASVAB. ARI is now programming EPAS to employ MPP as the objective function and also is developing a more valid predictor composite for attrition. They are also in the process of extending validities to more jobs through the synthetic validity approach referred to above. These studies should provide more precise estimates of utility and a better understanding of the interactions among parameters. But we fully expect the choice of the same alternative, optimal assignment using FLS composites and selection using FLS "g" composites (described in the next chapter) as being best for both selection and classification efficiency.

In closing this section, it is important to note that the performance gains forecasted as a result of the simulation results are realized only if there is better utilization of ASVAB information in operational practice. The recruit acquisition system must utilize FLS composites to the fullest, using optimal assignment algorithms, while meeting job quotas and quality goals, as assumed in our constrained FLS assignment strategy. Recruits would be required to accept assignments on their best or nearly best aptitude composite scores, rather than merely on the basis of their preferences and meeting low minimum composite cutting scores.

E. CHANGES IN POLICY BASED ON SIMULATION RESULTS

In this section we address the changes that could be recommended solely on the basis of our simulation results. In the final chapter, we will propose an integrated, phased sequence of recommendations based on research currently in progress, results of prior studies, and psychometric theory.

Only technical changes in assignment policy and procedures are necessary to obtain the productivity gains of the levels estimated in the present. The changes we propose call for the best use of all information contained in the present ASVAB along with a simultaneous increase in job standard minimum cut scores. Specifically, we suggest four

changes that appear to be implementable in the near term, provided the assumption and estimates made in our study hold in the specific decision-context of each service:

(1) As an immediate interim solution the raising of job standards, i.e., minimum cut scores, used for assignment to jobs, by at least five standard score units, rather than the use of the present low standards (until 4 below is implemented).

(2) The use of predicted performance (composites in standard score form times their validities) as the assignment variables, rather than the present practice of using aptitude area scores as the assignment variables.

(3) The use of full least squares prediction composites for each job family, rather than the use of unit-weighted, three-test aptitude area composites.

(4) The use of an efficient computer-based algorithm within EPAS, to optimally allocate personnel to maximize MPP, rather than the use of a system which makes virtually no use of AA composites or PP scores.

The assumption is made in these proposed changes that the preponderance of recruits can be persuaded to accept the jobs in which they perform best or nearly best. (This assumption is supported by past research and recruiting practices, particularly in the AF). Additionally we assume that the cost of implementing and operating a person-by-person, or sequential algorithm, for personnel allocation is nominal and needs no further technical innovations, especially considering the sophistication of systems now used in selection and classification by the services and those now under development.

Several decades ago it was essential that the calculations required to determine assignment to jobs be as simple as possible for operational use. To meet that need aptitude area (AA) composites of unit-weighted, three-test combinations of selected subsets of the total battery were developed. To achieve simplicity, since the current classification procedure uses one of nine AA composite scores to predict job performance in a job family, measures of *ability* (AA composites) were used in place of measures of *performance* (full LSE composites). Although it is widely recognized that the objective of a selection and assignment system is to maximize performance as a means of increasing utility, we believe that most researchers and policymakers were not aware of the extent that the current assignment system vitiated potential gains by replacing predicted performance measures with ability measures.

Modern computer-based person-job matching systems such as EPAS can feasibly provide for operational use of full least squares predictor equations--the optimally weighted ten tests of the ASVAB--for assigning enlistees to jobs. We noted in an earlier section of

this chapter that the use of LSEs would result in productivity gains more than five times greater than the gains of the current system over random assignment and simultaneously meet most of the significant policy and management goals. Appreciable gains would be realized by simultaneously raising job standards at least five points and possibly ten points even without the increased use of the recommendations of an optimal assignment algorithm. Since the current aptitude area system uses a reduced number of tests which are not optimally weighted in each AA, the composites are not best for classification as clearly demonstrated in our utility analysis (i.e., they needlessly reduce the PCE of the battery), just as the AFQT is not best for selection.

F. IMPLEMENTING THE SIMULATION RECOMMENDATIONS

In this section we examine the feasibility of implementing the recommendations in the near future, given the acceptance of the study's finding in the decision context of each service. We are not, however, recommending a decision on this set of recommendations based solely on the findings of the simulation. These findings should be incorporated into a broader set of other findings and conclusions discussed in the two following chapters.

The major implementation effort involves the development of an efficient computer-based algorithm for assigning individuals. Presently all the services use computer-based systems to accomplish various functions that facilitate the recruitment, selection and assignment processes. But the use of an operationally effective LSEs system would place special demands on the allocation procedure: the procedure must use a "line by line" linear programming algorithm in making recruit assignments that depend on the availability of accurate, complete and timely information.

As described in Chapter 2, the Army's EPAS system also requires information on recruit supply, training needs, and policies. To obtain such information, EPAS is building a number of interfaces with existing data bases; the information system and procedures necessary to maintain the system need to be in place before EPAS can be used operationally. The EPAS system uses this information to generate forecasts of recruit supply and training requirements that input to an optimization model. Optimization first ensures that all requirements, targets, and policies are met. Then among these feasible alternatives the optimization finds the distribution of recruits that maximizes performance and minimizes attrition from the pool of available recruits. Information on the optimal distribution of recruits is used to classify recruits on the basis of how they compare to the best available candidates who are likely to enlist.

The use of an LSEs strategy basically requires two modifications of the EPAS system: more frequent updating of the allocation plan (e.g., once every two weeks) and the addition of a "column constant" to each recruit's LSE score for each job. The column constants can be generated by the use of a model sampling technique that optimally distributes synthetic entities by procedures described in an earlier report (Johnson and Zeidner, 1989). These adjusted scores are then used by EPAS on a line-by-line basis in making person-job matching assignments. The essential distinction, then, between EPAS as currently designed and one that would use LSEs is that the latter places greater emphasis on using algorithms that optimize mean predicted performance.

LSE scores could be scaled so that they resemble traditional aptitude area scores. LSE scaled scores could be transformed to have a mean of 100 and a standard deviation of 20 multiplied by R, the multiple correlation for each job family. Thus the LSE scores can be used in the same manner as aptitude area scores in counseling recruits, specifying cut scores and in record keeping.

In the next several years, the results of several developmental and research efforts will be available that should provide further improvements in selection and classification. The Army's synthetic validation effort may permit extension of a high quality estimate of validity such as provided by Project A to 30-50 job sub-families. In assigning recruits, separate LSEs could be computed for each of these sub-families, rather than the nine job families now in use. The increase in the number of LSEs used will result in an increase in PCE capitalizing on: (1) decreased intercorrelations among FLS composites, (2) more opportunity for increased variance within persons across jobs, and (3) increased predictive accuracy due to the greater homogeneity of job families.

Also job value weights may be available for use in computing LSEs on the importance of various performance level/job combinations. Inclusion of such job value weights as assignment variables will increase the PCE of the battery by capitalizing on hierarchical layering effects.

Additionally, composites that differentially predict job performance and attrition based on a bro r range of tests and biographical information may be available for inclusion in a new ASVAB used in assignment. Increased differential validity of an improved ASVAB would increase PAE.

The possible changes noted above can readily be incorporated within EPAS; taken together they offer the potential of greatly increasing the utility of the assignment process.

The first three of the model sampling experiments described in the following chapter should be completed in the next year. The results may offer promise of even greater future changes in personnel utilization effectiveness.

We wish to emphasize that adoption of our recommendations for implementation in the future is based on sound theoretical, empirical and practical considerations. Our utility analysis, given the assumptions used, result in huge productivity gains over random selection and assignment.

In closing this section, we note again that no change in the present ASVAB is required to attain our estimate of productivity gain; the gain is attained by using available information more fully and rationally and employing an LP assignment algorithm that maximizes performance while meeting all constraints. The productivity gain is attainable without additional research to improve ASVAB validity or its differential validity, although the results of such ongoing research have the potential of greatly enhancing the utility of selection and classification in several years.

CHAPTER 5. NEW RESEARCH ON CLASSIFICATION EFFICIENCY

A. RESEARCH QUESTIONS PERTAINING TO THE ARMY CLASSIFICATION SYSTEM

1. An Approach to the Determination of Research Promise

The most important lesson to be learned from the simulation and utility analysis of Chapter 3 may well be: firstly, that small gains in MPP for an Army cohort group translate into hundreds of millions of dollars worth of increased productivity; secondly, attending to the psychometric principles of classification efficiency can yield such a gain; and, thirdly, an analysis of opportunity costs may show that the increased productivity obtainable from improving classification efficiency may cost many millions of dollars more if obtained by any other way.

There are a number of promising changes in the Army's operational selection and classification system, in addition to the substitution of FLS composites for the existing AA composites, that could provide appreciable amounts of productivity that, for the most part, are additive to the gains due to the use of FLS composites. These gains, like those confirmed in the simulation of Chapter 3, can be predicted on the basis of psychometric principles, and even to some extent by the unfortunately sparse prior research results. While one can be rather certain some improvement would result, the limitations of the prior studies, particularly their failure to report their results in utility terms, make it difficult to approximate the probable magnitude of gain obtainable from these promising operational changes.

We will look at three sources of information in considering several promising operational changes: (1) psychometric principles of personnel classification as reported in Johnson and Zeidner (1989); (2) prior research results; and (3) the Chapter 3 simulation and utility analyses. We will show where model sampling approaches can be applied to existing data to confirm that expected gains are of a magnitude that provides practical significance. We will describe model sampling experiments, ones we are initiating

concurrently with the writing of this report, as both "short-term" augmentations of prior results and as confirmations of gains we predict on the basis of psychometric principles. We believe these short term experiments will confirm our overall expectations as to the gains in productivity obtainable from our recommended changes in the operational system we describe in Chapter 6.

2. Four Promising Areas for Obtaining Increased PCE

A set of test composites can provide no more PCE for a prescribed set of job families than was provided in the test selection process that created the operational test battery. PCE can be increased for a fixed operational battery only by increasing the number of job families with associated test composites, and even then only if this shredding is accomplished in a classification efficient manner.

No improvement in the PCE of composites can result from removal of highly correlated tests from a set of FLS test composites or from the application of any other procedure for reducing the intercorrelations among FLS composites. The FLS composites already provide the maximum amount of PCE for a fixed battery and specified set of jobs or job families. However, improvement in PCE can come from selecting tests with high PCE for inclusion in the operational battery. This can be accomplished in the following two steps: (1) the selection of predictors which experts believe have a high degree of differential validity (as contrasted with predictive validity) for inclusion in the experimental test pool; and (2) test selection using indices that measure PCE to identify the operational battery with the best PCE. Harris (1987) showed that the use of Horst's index of differential validity was successful in increasing PCE. For a selection of 5 tests from a much larger set of tests in an experimental battery using the PCE-sensitive index, he demonstrated a 10 percent increase in MPP over that provided by 5 tests selected using an index which maximized predictive validity.

Given that a small number of FLS composites (from 9 to 12) are being used to assign to the same number of efficiently determined job families, a worthwhile improvement in MPP can be obtained by a major increase in the number of job families. An increase in the number of composites and associated families to somewhere between 20 and 40 would most likely provide the maximum efficiency for Army jobs, assuming that data are available from which to compute moderately stable FLS weights for the composite associated with each family.

Unless the current system is changed, the use of 20 to 40 separate test composites would require the Army to record this many scores on each soldier's form 20. One way to use this many assignment composites would be to install a two-tiered system in which the large number of FLS composites are used to make recommendations regarding assignment, while a much smaller number of factor scores are used for counseling. These factor scores would also be used as a basis for setting minimum cutting scores for entry into special training programs, as a career planning aid to be available to the soldier, and for other personnel management purposes, such as retention and promotion.

Our Chapter 3 simulation, along with the prior results of Harris (1967) and Sorenson (1965), provides convincing evidence that improving assignment procedures can produce dramatic increases in MPP. The largest and most dramatic increase in MPP will undoubtedly come from the use of FLS composites for both selection and classification in a two-stage selection/classification process. A smaller, but still worthwhile, improvement will result from the integration of selection and classification procedures using the MDS algorithm.

While prior results and psychometric principles virtually assure us that the use of a MDS algorithm will provide a dollar gain of practical magnitude, the estimate of this dollar amount must be more precisely determined before such a major policy change can be recommended to management. A model sampling experiment to provide this estimate has been initiated.

In summary, four promising operational changes are to be investigated in the first four of a series of model sampling experiments to be initiated in 1989:

a. Replace or augment existing ASVAB tests with new predictors selected from Project A experimental variables, using a test selection index which maximizes PCE rather than predictive validity.

b. Determine the PCE provided by 3-, 4-, or 5-factor scores, as compared to Army AAs and FLS composites; the factors on which scores are based will be obtained using an approach which maximizes PCE.

c. Determine gains in MPP obtainable from use of MDS.

d. Determine the upper bounds of gains obtainable from shredding selected Army job families into sub-families, then estimate gains in MPP obtainable from increasing the number of job families using an optimal clustering algorithm that maximizes PCE.

B. FOUR MODEL SAMPLING EXPERIMENTS

The four experiments we describe in this section correspond to the four objectives described above, except that each experiment will also investigate related psychometric issues pertaining to classification efficiency. For example, the first experiment is primarily concerned with selecting tests to maximize PCE, but it also investigates the effect of doubling the number of jobs to which individuals are being assigned--from 9 to 18--has on the magnitude of MPP. Since the 9 jobs represent 8 of the 9 Army job families, and all 18 fall into different "sub-families", the results from this aspect of the first experiment can immediately confirm (or dispute) the efficacy of shredding out the existing 9 job families into 18 families (based on the existing sub-family structure). The fourth experiment provides more precise information on the benefits obtainable from a more efficient approach to increasing the number of job families.

All four experiments assume the covariances and validities against a job criterion to be those of the youth population. These universe relationships are based on a Project A study in which scores for 29 predictor variables and 5 criterion components were collected for 19 MOSs (jobs). The experimental tests include measures of spatial visualization and orientation, perception and psychomotor skills, temperament/personality, vocational interest, and job orientation. Separate performance (criterion) scores are provided for 5 components: (1) specific MOS skills; (2) general (military) skills; (3) leadership behavior; (4) personal discipline; and, (5) military bearing/physical fitness.

A corrected matrix of covariances for 29 predictors, bordered by 19 validity vectors, is used to define the youth population. A two stage correction for restriction in range (selection effects) is applied to the empirical results of Project A to obtain this estimate of the youth population. First, a correction for the incidental effects of selection on 9 ASVAB tests for which youth population data are available was accomplished on all predictor covariances for the total sample of incumbents of all 19 jobs. This corrected total sample matrix is then used as the "universe" covariance matrix to correct validity vectors for direct selection effects in each job sample. The corrected validity vectors are thus made comparable to the already corrected covariance matrix for the total sample.

The synthetic scores for each sample will be generated by the model sampling procedure described in Johnson and Zeidner (1989). While the estimates of youth population relationships among the predictors, and between predictors and criterion variables provide the starting point of all the variables used in the four experiments, the experimental conditions are in part reflected in the choice of variables and weights used to

create assignment variables, and in part by the simulated selection and assignment procedures. Each artificial person, i.e., entity, will have assignment variable scores that reflect the experimental conditions before the simulation of personnel procedures begins.

1. First Experiment

The primary objective of the first experiment is to determine the gains in MPP that can be obtained using an algorithm for sequentially selecting tests to maximize a specified index. A 5-test and a 10-test "battery" will be selected using each of 5 indices: two of these indices have been proposed as a means of maximizing potential allocation efficiency (PAE); two others for maximizing PCE; and one for maximum predictive validity, i.e., potential selection efficiency (PSE). A total of 600 samples of 200 entities will be assigned under 30 experimental conditions.

This experiment will also compare the amount of MPP resulting when 200 entities are optimally assigned to 9 jobs as compared to 19 jobs. Brogden (1951) provides a table of mean criterion standard score values when all entities are assigned to up to 15 jobs (p. 190). If his tabled variable is called M, the MPP is equal to $M(R(1-r)^{1/2})$. Under Brogden's assumptions, increasing the number of jobs (or job families) from 9 to 15 provides an increase in MPP of 16 percent.

Brogden's assumptions include stipulating the same values for R (the average validity of FLS composites) and r (the average intercorrelation of FLS composites associated with each job). However, the value of R should be increased and r reduced by some unknown amount as job families are made more homogeneous. If R is increased from 0.70 to 0.75 and r reduced from 0.95 to 0.90, the MPP would be increased by 78 percent, i.e., from 0.23 to 0.41--if Brogden's other assumptions are met. We feel certain that his assumption that the covariance among PP variables can be explained by Spearman's "2 factor" theory is untenable. And, we do not know how robust Brogden's assumption might be with respect to this assumption.

Taking all of the above into account we believe there is a strong possibility that increasing the number of FLS composites from 9 to 15 would provide as much an increase in PAE as can be provided by improving the ASVAB on the basis of Project A results. It is unlikely that even an optimal reconstitution of the ASVAB by making deliberate selections from the Project A experimental test pool to maximize PCE would increase PCE more than what could be accomplished by increasing the number of job families from 9 to 15. This experiment is discussed further in Appendix 5.A.

2. Second Experiment

The dimensionality of the joint predictor-criterion space is probably no more than 4 or 5; if so, this space can still be used to identify one PSE efficient set of factors and one PCE efficient set of factors. As many as 30 FLS composites could be closely duplicated by linear functions of a general FLS composite and 4 or 5 FLS composites that define classification efficient factor scores. As the number of FLS job specific composites are increased, it becomes increasingly attractive to have a two-tiered classification system in which the larger number of FLS composites and procedures used to make initial assignments are essentially "black boxes" invisible to applicants and recruits, and the scores of a small number (5 or 6) of composites that define factor scores are placed in the recruit's official record and used for counseling and to make personnel management decisions.

Before making a decision to install a two-tiered system we believe policymakers should wish to know how much PCE can be provided by a small number of factor scores. Our second experiment is designed to compare the PCE provided by 12 different types of assignment composites, including FLS composites--the set which necessarily provides the maximum PCE--and the Army AA composites--the set which we fully expect to provide the least PCE.

The other 10 composites are based on one or the other of two kinds of factors. One approach provides factors that successively maximize the factor contributions for criterion variables. This is equivalent to maximizing Horst's (1954) "absolute validity" index for the first factor, then for the second factor in the residual space, and so on for as many factors as are extracted. The other type of factor, in a similar fashion, successively maximizes Horst's differential validity index over the total set of criterion variables.

Both types of factors are identified and rotated in the joint predictor-criterion space and then extended (Dwyer, 1937) to the predictor variables. All rotated factors are expressed as best weighted composites of the full set of predictor variables; factor scores can then be computed. Predicted performance scores are computed as linear functions of the factor scores. These functions range from those using only weights of 0 or 1 to those with weights that are signed integers of 1 through 4.

A cross validation design is used in these model sampling experiments. A sample of synthetic entities with the exact number of cases for each job--as were in the empirical data collection that provided our universe covariance-validity matrix--will be generated as

the first step in this experiment. The matrix of predictor covariances bordered by validities provided by this "random" sample has the statistical characteristics of being computed on an independent sample drawn from an infinitely large population, as defined by our universe matrix. This sample has the same number of entities contributing to the computation of each cell value as there were cases on which to compute the values in the uncorrected empirical sample. Thus we have a means of reflecting the effects of the sampling error that is necessarily present in the small to moderately large Project A job samples used to compute validities.

The parameters defining all assignment composites will be computed on the above sample and the weights for the FLS composites to be used to provide MPP scores computed on the sample designated (assumed to be) the population. These two independent sets of parameters--one for use in assignment and the other for evaluation of results--permit the avoidance of correlated error between the experimental assignment variables and the evaluation variables. Both sets of parameters will be used in several hundred cross-samples of approximately 200 entities each. For each of these cross-samples the PP scores will be computed using the parameters of "sample 1," the optimally-assigned entities constrained by the quotas on the basis of these PP scores, and the MPP standard score computed using the parameters of "sample 2" on the cross sample entities after assignment to jobs.

We are confident that FLS composites will not be shown to provide a significantly greater amount of MPP than can be provided by least squares estimates (LSEs) that use classification efficient factors as the independent variables. We feel only a little less certain that the composites of factor score using weights of signed integers will provide an adequate approximation to the PP provided by FLS composites. This experiment is required to confirm our predicted results, as well as to resolve doubts as to the feasibility of using the least complex of the proposed functions of factor scores as surrogate assignment variables. If these simplified surrogate functions provide a sufficient amount of PCE, one could defend the use of these surrogate functions by the counselor, by supervisors, and the individual in self-assessment. Appendix 5.B provides detailed information on this experiment.

3. Third Experiment

The third experiment focuses on selection and classification strategies. Three alternative strategies are evaluated under conditions (data characteristics) in which

hierarchical layering is present only in the general factor, only in the group factors, in both, in neither, or present in both to the same degree as is found in the Project A data. The three strategies are as follows: (1) selection on an FLS "general" composite in stage one, and optimal assignment to jobs using FLS job family-specific composites in stage two; (2) simultaneous selection and assignment using one PP measure--effecting pure hierarchical classification by multiplying the standardized scores of the FLS "general" composite by the validities for each job to create as many PP scores as there are jobs; and, (3) simultaneous selection and classification using the MDS algorithm with the same selection-classification variables as used in the two stages of the first strategy. The interaction of two levels of selection ratio with the effects of the five data characteristics and three strategies will be determined.

We propose immediate implementation of the first strategy in Chapter 6. The immediate operational implementation of the second strategy has been urged by several prominent investigators in the university sector. The third strategy is a highly promising approach whose benefits should be confirmed before implementation is seriously considered. Thus, we feel that the MDS approach is being compared with its two principal competitors in this model sampling experiment.

4. Fourth Experiment

The increased PCE that might be obtainable from reconstitution of Army job families will be considerably underestimated by the preliminary results provided by the "first experiment". The Project A data set provides 19 MOS that have been classified into separate sub-families within the present system. Since this family structure evolved primarily by expert judgment in which a number of considerations other than PCE had priority, there is a strong possibility that increasing the total number of existing job families could be accomplished more effectively than by adopting the current job family structure. This experiment will provide information on the benefits of using two alternative job clustering concepts, and will investigate the usefulness of using a data bank that contains test scores for the ASVAB tests and criteria for 98 Army jobs. These test and criterion data derived from 1981-82 Army accessions are described by McLaughlin et al. (1984).

The 19 jobs used in the above three experiments will provide the basis for comparing the conclusions reached from the more expensively obtained Project A data with the "81-82" data. The latter has many more jobs and larger N's for each job sample, but uses criterion data collected for purposes other than selection or classification research.

The reliability of the criterion measures tend to be lower probably because they were collected for evaluating the effectiveness of training programs and to implement a program for rewarding soldiers who meet minimum standards for their skill level. They are not sensitive at the upper range of productivity. This makes them appear to be less appropriate for use in our research than the criterion variables of Project A. However, if conclusions reached using the "81-82" data are essentially the same as those reached from use of Project A data, further analyses of the "81-82" and similar, more affordable, data will be justified.

This experiment will use a cross-validation design much like that of the "second experiment." The same analysis sample matrix used, respectively, to compute assignment variable parameters and evaluation parameters will be used for similar purposes in this experiment. The "analysis" matrix will be used to cluster the 19 jobs into 6, 9, and 12 families by each of two clustering methods and to compute the weights for the assignment variables. The population will again provide the weights for the FLS composites for computing MPP in each of the cross-validation samples.

The nine ASVAB test variables will be selected from the 29 variables of the "analysis" covariance validity-vector matrix to provide a 9 by 9 predictor covariance matrix bordered by 19 validity vectors (a 19 by 9 supermatrix). This 19 by 9 analysis matrix will be used to accomplish the clustering of jobs, separately by the two methods, into sets of 6, 9, and 12 families. A third and fourth structuring of the 19 jobs into sets of 6, 9, and 12 families will be identified using each of the two methods applied to the 19 by 9 covariance-validity vector matrix derived from the "81-82" data. Thus there will be a total of 12 sets of job families from which the covariance among the FLS composites corresponding to each job family will be computed.

The research design can be summarized as including the following main effects: (1) clustering methods (two levels); (2) source of predicted performance data on which to accomplish clustering (two levels); and, (3) size of job families (three levels). Thus there are 12 cells in our results matrix for which we will generate 30 cross-samples (replication) of 216 entities for each of these cells--calling for a total of 360 simulations. In each simulation 216 entities will be optimally assigned to 9, 12, or 16 jobs and the MPP score computed.

The results of this fourth experiment may encourage us to propose using the existing job family structure in the Army to create up to 30 job families, each with its own FLS composite. On the other hand, the results may suggest the desirability of conducting a

full fledged reconstitution of jobs into job families before adding more than a half dozen families and their associated test composites. We feel certain that this experiment will provide strong evidence against the reduction of Army composites from 9 to 4, as has been seriously proposed. We feel strongly that the trend in the other services to reduce the number of composites to correspond to the dimensionality of the joint predictor-criterion space is a mistake and should be reversed.

C. RESEARCH ON OPERATIONAL PROCEDURES

1. Recruiting and Counseling

We have provided evidence for the availability of greatly increased utility from the application of psychometric principles to the operational classification system. However, this increased potential utility becomes actual utility only to the extent that optimal, or near optimal, assignments can be enforced or sold to applicants. Since complete enforcement could work only with a draftee input we do not have, we should focus on selling the applicants on accepting an assignment which is available and for which his PP is comparatively high.

The first step in selling an applicant is to provide an ordered list, sequentially brought to view, catering to the applicant's needs, rather than reflecting the needs and convenience of the acquisition and training system. An efficient and equitable ordered list must comply with a number of principles, including the following three: first, we should always ensure that the information used to sell the ordered list is truthful and relevant to the applicant's long range career objectives; second, utility must be served by providing more leverage to the superior applicant. An applicant who would raise the mean MPP score of the job incumbents should have his preferences given more consideration than an incumbent who would lower the MPP score if given his preference; and, third, the process that produces the ordered list must appear fair to applicants competing for a limited number of seats in a training program. An applicant who ranks high among his competitors should not lose out because he has even higher PP scores for other jobs that he likes less. All three of these principles work against achieving all the benefits of a high PCE that could be achieved under enforced optimal assignment. Nevertheless, we believe an improved presentation of information to the applicant based on knowledge of the applicant's abilities and goals, combined with a deliberate effort to achieve classification efficiency could bring about most of the advantages obtainab e from optimal assignment.

We realize that such promising techniques as person-by-person assignment, and simultaneous selection and assignment using the MDS algorithm, need to have their impact on recruiting and assignment counseling determined prior to actual implementation. It is undoubtedly essential to combine the use of a person-by-person assignment capability and a revised "ordered list" with the MDS algorithm before the MDS can be considered to be operationally feasible.

In the absence of an optimal classification algorithm in the Army system for recruiting, classification, and the making of initial geographical assignments, the usefulness of the AA composites is dependent on the effectiveness of minimum cutting scores. There would have been literally no effective use of the ASVAB by the Army during the past two decades if the use of cutting scores did not exist. Thus the lowering of cutting scores to the point where almost all recruits are eligible for all jobs resulted in a mere token use of the ASVAB, except for selection where only the AFQT is used.

For a classification system in which assignments are made with an LP algorithm, the use of such cutting scores could only lower MPP, the objective function. Thus, the Army has the option of (1) continuing a token use of AA composites or even eliminating the use of more than one composite, retaining only a single measure for selection purposes, (2) improving the quality of minimum cutting scores, controlling the assignment into jobs, thus restoring the usefulness of the AA composites, or (3) installing an optimal assignment algorithm into the system and using it to effect assignments.

The simulation described in Chapter 3 raised the existing official minimum cutting scores by 5 Army standard score points (one-fourth of a SD). The assumptions relating to costs associated with the additional recruiting that use of cutting scores would require were very conservative. The closest simulation of a preference driven assignment system is provided by the current system model with resulting effects of higher cutting scores estimated by a credible (to us at least) job-person match procedure and reasonable pliant recruit behavior. The resulting gains in MPP are considerable, but so are the estimated costs for recruiting replacements. When costs are subtracted from gains in predicted performance, it becomes apparent that the utility of the changes in magnitude of minimum cutting scores depends on the recruiting strategy, with its associated costs, that one believes will be adopted to replace those excluded by the raised cutting scores.

More utility could be obtained from use of minimum cutting scores computed to be proportional to the column constants provided by a dual LP algorithm. In a hypothetical

example in which all recruits are willing to choose again whenever they fail the cutting score of the job they chose, and preferences are perfectly correlated with ability, MPP would be maximized--the same assignments would occur as would be made from an LP algorithm that maximizes predicted performance. This example is, of course, unrealistic, but we do not know how unrealistic it really is. Research to provide data on relationships between preferences and aptitudes, and among preferences, accuracy of the applicants' knowledge of Army jobs, and willingness of the applicant to accept alternative jobs, becomes highly desirable. With this data, an accurate simulation of the recruiting assignment process can be provided, and the utility of alternative sets of minimum cutting scores compared.

Several decades ago, when ARI provided minimum cutting scores for entry into all Army school courses, these scores were objectively computed as a function of the average AA composite scores of students entering all courses, the percent failing, and the validity of the AA composite against final course grade. The score at which 50 percent or more of the students were predicted to be graduates was selected as the cutting score. As the training philosophy was changed to produce few if any failures, this algorithm became obsolete.

Cutting scores are now negotiated on the basis of a number of factors. The difficulty of school courses, criticality of errors on the job, "value" of the incumbent's product, and recruiting problems are all considered, at least informally, and integrated to create the cutting score for an MOS. The quality distributions provided as recruiting and assignment goals are also subjectively determined and have certain similarities to the cutting scores in that both reflect the value, as well as the difficulty, experts estimate for the different jobs. The latter determination differs from that for cutting scores in that the former relates to AA composites and is expressed as a single score, while the latter relates to AFQT, more of a measure of general mental aptitude, and is expressed in terms of the minimum number of recruits desired to fall into each of the 4 categories of AFQT (I, II, IIIa, and IIIb).

With very low minimum cutting scores on the AA composites there is very little conflict between meeting quality goals and complying with minimum cutting scores. If cutting scores are selectively raised for some jobs, and not others, the pattern of MPP scores would also be altered and the meeting of quality goals might not be so easily attained. An initial impact study of this sort could be readily accomplished using either the simulation approach demonstrated in Chapter 3 or model sampling as described above. However, the additional data relating performance, ability, and willingness of the applicant

to continue negotiating when the first choice is not available, is required before a precise impact study can be accomplished.

2. Reconstitution of Job Families

The current Army job families and sub-families represent career ladders and reflect both the curriculum structure and the responsibilities for on-the-job training of Army schools. It is not necessary to disturb this deeply entrenched infrastructure in order to use different job clusters for the sole purpose of providing more effective FLS test composites and their corresponding job clusters for initial assignment. However, it becomes even more clear that our proposed two-tiered system is an essential ingredient of any system change which either involves a moderately large increase in numbers of composites, or calls for a reconstitution of the existing job families into clusters to be used only for making initial assignments. In a two-tiered system, the smaller number of visible factor scores used for minimum cutting scores and all other personnel management practices except initial assignment, could be used with an unchanged job structure while the FLS composites could be related to job clusters used only for initial assignment.

There is an obvious need for each service to conduct its own research on job clustering, including a determination of utility gains obtainable from using more composites, the identification of the additional FLS composites, and determining the impact of a two-tiered system would have on its personnel management system. The Air Force would appear to have the most to gain since they now use only four AA composites, but they already have half of a two-tiered system in place--the visible part--and need only replace their four AA composites with many more FLS composites in the software of their assignment system.

Research should be conducted in the areas where the greatest opportunities exist to obtain utility gains. The most promising opportunity, second only to the introduction of FLS composites and the use of predicted performance instead of aptitudes, is the use of any increased number of composites for initial assignment. Personnel researchers and management analysts in all services must provide additional service-specific information before assignment systems capable of realizing these gains can be implemented. Unfortunately, during the past decade, the services have had little interest in conducting these kinds of research and management studies.

D. RESEARCH ON THE CONTENT OF THE BATTERY

1. The Issue of Dimensionality

The dimensionality of the joint predictor-criterion space is a major limiting factor to the increase of the PCE provided by the ASVAB. Dimensionality is what has to be increased if substantial increases in PCE are to be obtained for future batteries. We define a "dimensionality of n" as present when a statistical test will reject the hypothesis that (n – 1) FLS composites can explain the relationship among PP scores in an empirical sample that is independent of the samples on which weights for the FLS composites and PP scores were computed--and the same hypothesis with respect to n such FLS composites cannot be rejected. The joint predictor-criterion space is defined in terms of a specified predictor battery providing PP variables for a specified set of jobs. A factorization of the covariances of the PP measures is descriptive of their joint space, but determination of dimensionality also requires a test of significance.

Some of the more avid validity generalization advocates would argue that the dimensionality of the "joint" space is 1, that there is only a single measure that provides either significant prediction of criteria or PCE across jobs. Others would argue for at most three such dimensions: (1) a general ability factor; (2) a "psychomotor" factor; and, (3) a "speed" factor. Others would add one more motivational-interest factor and still others would add one or more technical information factors. Thus, there is serious advocacy for dimensionality for the "joint" space ranging from one to about a dozen.

A question as to the dimensionality of the "joint" space requires an answer based on the examination of empirical data using the scientific method. In contrast to simulation-utility studies where the magnitude and value of utility gains are of primary interest, the traditional scientific method requires the statement of a null hypothesis plus an alternate hypothesis and the use of a statistical test that permits either the rejection of the null hypothesis or the failure to reject it. The "alternate" hypothesis is, to varying degrees, implied as at least a possible explanation. This "alternate" hypothesis cannot be proved, but only "accepted" with a meaning hinging on several other considerations, such as the extent the law of parsimony or Occam's razor is fulfilled.

For some simulation purposes the empirically obtained matrix of covariances among PP scores can be utilized as descriptive of the universe from which samples can be appropriately drawn. Such a matrix, after a few almost trivial adjustments, is our best estimate of the population covariances. If based on at least a moderately large sample, this

matrix provides a credible representation of the population from which the sample was drawn.

However, if the investigation has hypothesized that the PP variates are comprised of only one measure plus error, it would be permissible to test statistically the hypothesis that a single measure could explain the empirical data. If these tests fail to reject the hypothesis of unidimensionality, no one has proven anything about dimensionality, but the investigator would be justified in "accepting" the unidimensional hypothesis. He would then be justified in using a maximum likelihood factor solution to reproduce a covariance matrix in which each covariance value reflects an underlying factor. This matrix might then be used to generate samples of synthetic scores for use in a model sampling simulation--one in which the assumption of unidimensionality is explicitly made.

The above mentioned example was provided to illustrate how research on the dimensionality of sets of PP variables, as present in the real world, could appropriately constitute a preliminary step to the conduct of simulations for the evaluation of operational systems. We will next show how Wise et al. (1987) used the scientific method to investigate the dimensionality of a set of PP scores provided by the Project A data. They used a statistical package, LISREL, to test the hypothesis that a single measure, equivalent to what we refer to as the FLS "g" composite, could explain the covariances among the FLS estimates of performance for a small set of Army jobs on which Project A data was obtained. The rejection of the hypothesis that the variances and covariances of the PP estimates can be explained by a single measure that most investigators would be willing to call general mental ability raises the question of how many FLS composites can be hypothesized and still obtain a rejection of the null hypothesis.

Maximum likelihood factor analyses and multivariate statistical tests can be used to determine whether the hypothesis that the sample represented by a covariance matrix was drawn from a specified population as represented by another covariance matrix. We do not discourage the use of such tests. However, we commend the much easier to understand cross-validation design, including confirmatory factor analyses, to investigate whether "n" FLS composites, where weights are defined in an analysis or "back" sample, can be confirmed as providing the benefits of multidimensionality in independent, "cross" samples.

As a notional example to illustrate a method which we will later propose as a way to select tests, assume that three jobs or job families are selected, through analysis of prior

results, as promising candidates for having distinguishable PP composites. The weights for defining these composites are computed in analysis samples A_1, B_1, and C_1--to define FLS composites *a, b, c,* respectively. Each of the two possible comparisons of validities for these composites are examined in each of the cross samples A_2, B_2, and C_3. To reject the hypothesis that any pair of the composites lack statistical independence, two validity differences: $(r_{ay} - r_{by})$ and $(r_{ay} - r_{cy})$ are tested in sample A_2, and two similar differences tested in samples B_2 and C_2. All 6 validity differences must be significant in order to reject the hypothesis that two dimensions are adequate to explain the data--i.e., to confirm the existence of three or more independent dimensions. A tentative dimensionality of three stands until a hypothesis that the dimensionality is three can be rejected in favor of a higher dimensionality.

One method for testing for the statistical significance of a difference between two validity coefficients calls for first making an *r* to *z* transformation and then computing critical ratios. We find that for a credible set of values: $r_{ab} = 0.85$, $1/2\ (r_{ay} + r_{by}) = 0.50$, and $N = 327$, a difference of 0.05 would achieve a .01 level of significance for a one-tailed test. Considering that for our notional example the achievement of significant differences are required in three independent samples to reject the null hypothesis, a smaller level of significance obtained in each of the 3 samples could combine to provide a 0.01 level of significance overall.

We are not proposing a particular design for investigating dimensionality. Instead we are recommending that at least one of the several available methods be applied to the many sets of data in the services appropriate for the study of dimensionality.

The present lack of attention to the dimensionality issue goes hand-in-hand with the research community's overriding emphasis on predictive validity and the general inattention to the improvement of the classification efficiency of operational procedures. Research evidence supporting a respectable dimensionality of the joint predictor-criterion space would undoubtedly result in more attention to practical ways of improving classification efficiency.

2. Selecting Measures for an Experimental Predictor Pool

It is important to consider PCE in selecting measures from an experimental predictor pool to replace the least effective of the ASVAB tests. The consideration of PCE is just as important when decisions are being made as to the membership of the experimental predictor pool to be used in a major study on which many research dollars are

to be expended. Only a few opportunities to assess PCE across a moderately large number of jobs (19 jobs in the case of Project A) can be anticipated in each generation. The most should be made of each such effort by making a preliminary selection of tests using more affordable research designs characterized by: (1) fewer "job samples" for each set of candidate measures; and (2) "concurrent" rather than "longitudinal" research approaches.

In developing tests for the experimental pool, every effort should be made to avoid constructs that rationally predict performance equally in all jobs. To this end, expert judges should be asked to identify (1) abilities that one sub-family of jobs needs more than others, and conversely, (2) abilities that some, but not all, sub-families do not need more of than a small minimal amount for effective performance.

To improve PCE it is not enough to add additional content domains. The addition of non-cognitive measures to a previously 100 percent cognitive battery would not necessarily increase PCE. A strong "g" factor exists in the non-cognitive domain just as in the cognitive domain. A measure of propensity to adjust to the Army can add to the predictive validity of the PP composite for all Army jobs. Were there data to show otherwise, we would have suspicions regarding the quality of the experimental test administration or criterion measures--or, for small samples, sampling error. Such a general non-cognitive measure would be more appropriately included as a member of the FLS "g" composite and used for selection.

E. RESEARCH ON UTILITY MEASURES

We have previously emphasized the distinction between a performance measure, as commonly used as a criterion variable in the conduct of personnel research, and a benefits variable which denotes, separately for each job, the contribution to the Army's mission provided by each level of the performance measure. Conversation of performance to benefits can be accomplished in many alternative ways. Some assume linearity between performance and benefits and others do not. Also, some possible approaches give priority to capturing the presumed non-linear relationship between performance and aptitude and "contribution," possible at the expense of inadequately determining the relative magnitudes of the contributions of each job. Other approaches would emphasize the determination of the contribution each job makes to a common metric that can be used to express value.

Thus far, we have not adequately portrayed the difficulties inherent in the creation of a value metric that is based on: (1) each job's contribution to the mission; (2) the relationship of predicted performance to this conribution; and (3) the use of equivalent

measurement units across all jobs. The information provided by prior research does not adequately permit an informed recommendation as to which procedure should be used for effecting a conversion of performance to benefits. Related problems must be solved to depict accurately the effect such a conversion would have on (1) the utility resulting from our other proposed changes in the selection-classification system and (2) the impact the use of value weighting of jobs would have on other personnel systems.

In our simulation we used performance, rather than benefits, as the measure that we aggregated across jobs and then converted performance into dollars. The use of benefits was not an option because we had no reliable information on the values policymakers place on the productivity of different jobs. In effect we were assuming: (1) equal values for productivity in each job; (2) a linear relationship between performance and productivity; and (3) a linear relationship between productivity and the value of this productivity expressed in dollars. It will be necessary to continue to make these same assumptions in subsequent utility analyses and interpretations based on the results of the confirmatory simulations we have initiated. Considerable research on job values should be completed and the reactions of policymakers obtained before the use of job values in such analyses can be justified.

The Army is already implying disparate values for Army jobs by a number of policies now in effect. Through the years, various policies requiring the distribution of personnel "quality" somewhat equally across jobs have been expressed in a number of ways. In the era just preceding the first use of a LP program to make initial assignments for a predominantly conscripted input, Pentagon sorters were used to equalize the percentage of college graduates provided to the combat arms and the technical services. It has been suggested more recently that equal numbers of categories IIIb and IV input should be distributed to each job family, whether combat arms or the technical services.

Quality distribution goals are clearly expressions of job values. The minimum cutting score designated for each job partly reflects an estimate of difficulty, but also a large measure of the job's importance and criticality--an alternate definition of value. The establishing of "perks" for those with command responsibilities (e.g., for soldiers with green shoulder tabs in the Army, and for (WW II) Navy NCOs, the right arm rates as contrasted with left arm rates) also expresses a concept of disparate job values apart from pay grades. The concept of varying values across jobs cannot be foreign to policymakers.

Traditionally, for good reasons, the criterion variables used as surrogates for a benefit measure were underlying normal distribution usually assumed. Among those wishing to deviate from the traditional approach, there has been more interest in a non-linear translation of a normative criterion to a non-normal distribution of value scores, than in determining either the relative value levels across jobs or the value weights for use in converting predicted performance scores differently across jobs.

Nord and White (1988) demonstrated a technique for transforming a job performance measure expressed as percentile scores into normative, non-normal, benefit scores. They did not consider differences across jobs of various incumbent population characteristics that are known to affect level of performance, such as (1) MPP, (2) average experience, (3) average civilian education, and (4) the relationship between experience and PP for the particular type of job. If the scores resulting from the non-linear conversion of rank-ordered PP scores into benefits were used as assignment variables in LP algorithms, we would expect some changes in the objective function, as compared with the use of the original PP scores. However, this change could be entirely downward, in contrast to the use of benefit scores which have also been adjusted for the value attached to each job as assignment variables. A considerable increase in the objective function would be expected if value weights were to be applied to both assignment and the benefit scores obtained by the method of Nord and White.

It is conceivable that information on the relationship between experience and performance could be obtained. Then by also using the already known relationship between PP scores and performance, correcting the benefit scores distribution (obtained by the method Nord and White) could be corrected back to a youth population. When faced with the same population with respect to the most relevant variables, it would become more permissible to treat the benefit scores as if they formed a common metric across jobs.

Our purpose is not to propose a particular research program, but to urge that a research effort on the value of jobs be initiated. That research should develop a technique for producing job values based on principles that are both acceptable to policymakers and credible to researchers. The principles should then be applied to a set of jobs selected by policymakers and the results shown to policymakers for their review.

If these first two steps succeed, the third step is to conduct an impact analysis of the use of job values. The research community should provide a simulation to determine the effect the use of job values has on MPP, the utility provided by the selection and

classification system, and the distribution of quality. Management analysts should investigate the effect the use of job values has on related personnel distribution procedures, on meeting recruiting goals, and on understanding and controlling a variety of cost factors. Only then should policymakers make decisions concerning the final design and implementation of an operational systems that uses disparate values for jobs.

APPENDIX 5.A

MAXIMIZING POTENTIAL CLASSIFICATION EFFICIENCY IN THE ASVAB: SEARCH FOR MULTIDIMENSIONALITY IN THE JOINT PREDICTOR-CRITERION SPACE

APPENDIX 5.A
MAXIMIZING POTENTIAL CLASSIFICATION EFFICIENCY IN THE ASVAB: SEARCH FOR MULTIDIMENSIONALITY IN THE JOINT PREDICTOR-CRITERION SPACE

Overview: A model sampling experiment utilizes parameter values obtained from a large empirical study (Project A) to compare alternative test selection methods when the objective is to improve PCE or PAE. The best index to use in test selection procedures for maximizing predictive validity against performance on several jobs (Max-PSE) is compared with two indices that estimate PAE (modifications of H_d and PDI), and two indices that estimate PCE (unmodified H_d and PDI). Two sets of tests (5 and 10 in number) are selected, each index being stored in a sequential, accretion type algorithm.

Synthetic score vectors (entities) are generated using the parameters associated with each set of tests selected using one of the indices, and with their validities for either 9 or 18 jobs. The predicted performance covariance matrices based on the selected tests are the expected values for the covariances of each entity sample.

Entities in each sample are optimally assigned to jobs, using both equal and non-equal variances for the assignment variables, and the MPP standard scores resulting from each assignment process computed. Thus both PAE and PCE values are provided for each set of entities.

Three conditions that affect the dimensionality of joint predictor-criterion space can be separately tested as main effects and in terms of their interactions with the 5 indices in affecting PAE and PCE. The primary interest, however, is in the main effects (in terms of PAE and PCE) provided by type of index used in the selection process.

Problem: Research on the development of new predictors for the ASVAB has been dominated by consideration of predictive validity, as contrasted with concern for the increase of potential classification efficiency (PCE). This emphasis on predictive validity has dominated both in the selection of tests for inclusion in the experimental test pool and in the selection of tests from the experimental pool for inclusion in the operational battery

Some researchers contend that PCE will take care of itself if test selection is based on predictive validity. Still others contend that the dimensionality of the joint predictor-criterion space is essentially unidimensional; almost all of the gains in MPP resulting from assignments are then attributable to hierarchical layering effects. The two indices with sensitivity to hierarchical layering removed can result in test sets with superior PCE only if the joint predictor-criterion space is multidimensional and the tests that maximize Max-PSE are not also the tests that maximize PAE.

The verification of the hypothesized superiority of the point distance index (PDI) over Horst's differential validity index (H_d) requires either a simulation (using a rich data base) or a model sampling experiment for its verification. PDI can be shown to be superior under certain contrived data conditions but it is not known how likely these conditions are to occur, or whether differences in the result from using one index instead of the other is of a magnitude large enough to warrant the attention of researchers.

An analytical proof of the superiority of PDI would require the solution of definite integrals whose reduction exceeds the present state of the art in mathematics. Thus a simulation or model sampling approach provides the only feasible ways of investigating the problems described above. The principal disadvantage of reliance on simulation or model sampling is the specificity of the findings to a particular situation (in this case to a situation defined by the parameters derived from Project A data). It would be possible to perturb the existing data or to make systematic changes in the parameters along theoretically pertinent lines as means of providing more general results. However, such a sensitivity analysis is not within the scope of this study.

H_d can be shown to be proportional to the square of PAE under certain assumptions, including one that indicates the absence of hierarchical layering effects; a considerably lower relationship of H_d with PAE exists when much of the magnitude of H_d is due to hierarchical layering effects. Thus, the reduction of the component of H_d due to hierarchical layering may increase the relationship of H_d and PAE. However, the payoff measure, the MPP standard score, may reflect major hierarchical layering effects (and most definitely does in the data used for this study). Thus, the benefits from using indices for test selection that have been corrected to make them insensitive to hierarchical layering needs to be investigated empirically, using either simulation or model sampling methodology.

It is believed that the two PDI indices, one for maximizing PAE and the other for maximizing PCE, are superior to the two H_d indices. However, a practical degree of superiority has not yet been established and H_d-based results are needed to provide a basis of comparison with the results of Harris (1967) who used H_d as his means of improving PCE.

The two indices modified to remove their sensitivity to hierarchical layering effects are expected to provide higher PAE than the unmodified indexes. Because of the reduced accuracy of the unmodified indicies as predictors of PCE that is introduced by the presence of hierarchical layering effects, the modified indices may also provide higher PCE.

Research Questions: The research questions relating to PCE in the joint multidimensional space are those that: (1) relate to the utility obtainable from the use of five alternative test selection indices under specified conditions, or (2) pertain to the correlation of either PCE or PAE to the five indices. The utility questions refer to the MPP standard scores after optimal assignment to jobs; these scores are measures of PAE if all the assignment variables have equal variances, and otherwise are measures of PCE.

We believe we could, within present knowledge, rank order the expected values of PCE and PAE that, under the various conditions defined in this study, would result from an infinitely large number of observations. Thus, we are primarily trying to determine if practical differences (that are also statistically significant) result from the use of one index instead of another under specified conditions. We are uncertain as to which pairs of indices will yield differences with practical significance from a utility point of view.

PCE/PAE Obtainable Using Alternative Indices

The questions relating to utility are:

1. Is there a statistically significant difference between the results provided by the "best" five and the "best" ten tests? If one can conclude that there is no difference, the data for these two conditions will be combined for the remainder of the analyses. We anticipate that this difference will not be statistically significant; if otherwise, the questions raised below will have to be asked separately for the two sizes of operational batteries (5 and 10 tests).
2. Is the difference in PCE and PAE resulting from the use of a PCE oriented index (H_d or PDI) as compared to a PSE oriented index (Max-PSE) of statistical and practical significance? This represents our primary research hypothesis, and will be first tested by combining conditions of: (1) number of jobs, (2) source of criterion components, ignoring the distinction between H_d

and PDI with respect to PCE and between modified H_d and Modified PDI with respect to PAE. We expect these differences to be statistically significant; if so, the differences in PCE and PAE will be separately examined for the PDI and H_d based samples, and then the effects of those two indices will be further examined under the conditions of number of jobs and source of criterion components. The extrapolation of results to the Army's population of jobs will be made on the basis of the influence of these conditions.

3. Is there a positive advantage, a practical and statistically significant difference, in the use of the modified indices (modified H_d and modified PDI) over the unmodified indices in the measurement of PAE?

4a. What effect does the increase in the number of jobs representing the population of jobs have on the magnitude of PCE and PAE? Is there a practical and statistically significant increase in PCE/PAE when V_{18} is used to obtain a 18 by 18 predicted performance covariance matrix, which is in turn used to provide the parameters for generating the entities used to obtain PCE/PAE values for each sample, as contrasted with the use of V_9 for the same process? We anticipate a significant increase in PCE/PAE associated with the increase in number of jobs, but we expect to find a quite different relationship between number of jobs and PAE than would be expected on the basis of Brogden's model (1959); for m = 2 to m = 9 when R and r are invariant we expect less of an increase than could be anticipated if each job provided a measurable increase in the joint predictor-criterion space (Brogden, 1959). But for complex reasons, including the probable increase in R and decrease in r, we expect a slower approach to the asymptote than Brogden's model shows. This increase in PCE/PAE, if any, will be used to extrapolate the results of this study to a larger population of jobs. The answer to this question has considerable theoretical importance to the designers of an approach for developing and using a system of synthetic validities for a population of Army jobs.

4b. Is there an advantage in having objectively rather than subjectively measured criterion components in achieving PCE/PAE; i.e., is there a practical and statistically significant difference in the PCE/PAE provided by use of jobs with higher quality performance measures in the model sampling process detailed in 4a? We suspect that a statistically significant difference would result with sufficiently large Ns, whether or not the difference has practical significance with respect to utility. The magnitude of this advantage is an important factor in the extrapolation of results to the population of Army jobs.

The Prediction of PCE and PAE

Approach: This study is based on a model sampling experiment; the experimental results are further interpreted in terms of utility. A youth population predictor

intercorrelation matrix, R_t, and three validity matrices (one for the 9 jobs that have validities based on objectively measured criterion components, $V_{9,o}$, one for those 9 jobs using subjective components in the criterion variable, $V_{9,o}$ and one for those 9 jobs plus 9 more that have only subjective components in their criterion variables V_{18}) provide the parameters for generating the synthetic score vectors comprising the entities (i.e., artificial people). The results of the experiment are in terms of MPP standard scores resulting from assignment of entities to jobs. These scores are a measure of potential classification efficiency (PCE) or of potential allocation efficiency, depending on whether LSEs or LSEs divided by the R_i (a validity, the multiple correlation coefficient beteen the predictors and the criterion for the i^{th} job) are used as the assignment variables.

Predictor tests will be sequentially selected, separately for $V_{9,o}$, $V_{9,s}$, and V_{18}, using five alternative indices to be maximized by the test selection process. These five indices are as follows: Max-PSE, H_d, H_d modified to eliminate sensitivity to hierarchical layering effects, PDI, and PDI modified to eliminate sensitivity to hierarchical layering effects. Thus sequential test selection will be accomplished ten times; ten separate test selection sequences ranging from one to ten will be determined. The "best" 5 and 10 test sets will be selected for each of the 5 indices, separately for the 9 and the 18 job sets.

A number of separate 9 by 9 and 18 by 18 predicted performance covariance matrices computed for each selected set of tests will be factored to provide transformation matrices used to compute score vectors. A total of ten such 9 by 9 matrices (5 using the objective criterion and 5 using only the subjective criterion) and another five 18 by 18 matrices, $C = V(R^{-1})V'$, will be computed. Separate sets of entities with expected covariance matrices equal to these predicted performance matrices will be generated and the entities assigned to jobs using a LP program. Two assignment procedures will be used for each sample of entities: one in which the LSEs with standard deviations of R_i are used as assignment variables and a second in which the assignment variables are the LSEs divided by R_i.

MPP standard scores will be obtained after the entities in each sample are optimally assigned to jobs; assignments are made two ways, once with, and once without, equal variances among assignment variables. These MPP scores are the units of analysis for the determination of the comparative effectiveness of the five test selection indices under two levels for each of three conditions that affect the dimensionality of the joint predictor-criterion space. These three conditions are as follows:

(a) number of tests selected to represent the joint predictor-criterion space (5 or 10).

(b) number of jobs used to represent the joint predictor-criterion space (9 or 18).

(c) Source of components in the criterion variable; dimensionality of the criterion as affected by source of criterion components for each job (objective or subjective sources).

Research Design: The model sampling process proceeds from the initial generation of two vectors of random numbers, one a 5- and the other a 10-element vector. Each of these random number vectors is transformed into 10 different row vectors of synthetic scores (1 by 5 and 1 by 10 matrices, respectively), each of which has an expected covariance matrix equal to the covariances among one of the sets of selected tests. Using an entity sample size of 200 (N = 200), each of the score matrices becomes either a 200 by 5 or a 200 by 10 matrix. We call this matrix Y. When each element of Y is divided by the square root of N to produce Y, $E(Y'Y) = R_t$, where R_t is, in turn, the correlation among each of the 10 sets of selected tests.

A matrix W can be readily computed ($W = R_t^{-1}V'$) that will yield the equation $YW = Z$, where Y is the N by n matrix just described, w is an n by m matrix, and Z is an N by m matrix of job criterion scores. C is the m by m matrix of covariances (variances in the diagonals) among the predicted job criterion variables (predicted performance measures). Defining $\mathbf{Z}$ as the matrix with each element divided by the square root of N, we can write: $E(\mathbf{Z}'\mathbf{Z}) = C$.

Each Z represents the m predicted performance scores of n simulated persons. An optimal personnel assignment algorithm is to be used to match each entity to a job and the mean predicted performance (MPP) standard score computed, assuming each entity to be optimally assigned. Two assignment algorithms (two separate entity/job match processes) will be utilized for each Z. One will assign entities using the predicted performance scores modified to have equal variances across jobs as the assignment variables. The second process will use the unmodified LSEs as the assignment variables. The MPP standard scores resulting from the first process provide a measure of PAE and the second process provides a measure of PCE. These values of PAE and PCE will be entered into the matrix of results and constitute the unit of analysis for all further statistical analyses.

Thus one replication of the model sampling experiment produces 15 separate PCE and PAE values based on a single N by n ($n = 5$ and $n = 10$) matrix of random numbers. The 15 separate predicted performance covariance matrices from which the model sampling

parameters for each replication are obtained represent the 5 different test selection indices and the three different sets of job criterion: 9 jobs with subjective criterion components, 18 jobs with subjective criterion components, and 9 jobs with objective criterion components.

A traditional factorial design using two levels of a "number of jobs" factor and two levels of a "source of criterion components" factor is unavailable to us since the Project A study did not collect data on objective criterion components for 19 jobs. Instead, we can only contrast results based on $V_{9,s}$ with those based on $V_{9,o}$ to test the null hypothesis that no difference in PCE/PAE results from the use of objective criterion components. And we can contrast the same results based on $V_{9,s}$ with those based on V_{18} to test the null hypothesis that no difference in PCE/PAE results from increasing the number of jobs.

The matrix of results has 15 cells associated with the parameters based on 5 selected tests and 15 more cells based on 10 selected tests. A total of 30 cells have 20 replications each. Thus 600 separate N by n Z matrices are generated and separate values for PCE and PAE computed. The matrix of results in terms of factors and cells are provided in Table 5.A.1.

Table 5.A.1. Matrix of Results (Separately for PCE and PAE)

<table>
<tr><td colspan="2" rowspan="2">Predictor Related Conditions Affecting MPP
Further Conditions That Affect MPP</td><td>Test Number</td><td colspan="5">5</td><td colspan="5">10</td></tr>
<tr><td rowspan="2">Indico Used to Select Tests</td><td rowspan="2">H_d</td><td rowspan="2">PDI</td><td rowspan="2">Mod. H_d</td><td rowspan="2">Mod. PDI</td><td rowspan="2">Max-PSE</td><td rowspan="2">H_d</td><td rowspan="2">PDI</td><td rowspan="2">Mod. H_d</td><td rowspan="2">Mod. PDI</td><td rowspan="2">Max-PSE</td></tr>
<tr><td>Number of Jobs</td><td>Source of Criterion Components[a]</td></tr>
<tr><td rowspan="2">9</td><td colspan="2">Including hands-on measures</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="2">Does not include hands-on measures</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>18</td><td colspan="2">9 jobs with and 9 jobs without hands-on measures</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

NOTE:
a It is believed that the dimensionality of the joint predictor-criterion space is increased by the presence of hands-on measure of performance.

The use of the same random number vector for transformation into a row vector for each of the N by n Z matrices associated with one of the 15 cells noted above reduces the proportion of error variance found in the differences between cell means. Thus a repeated measure design can be used and fewer replications are required to assure that a difference large enough to have practical significance will have statistical significance.

The research questions mostly pertain to differences between cells rather than among levels in a factor. Thus, the factorial design is intended to provide a preliminary justification for the examination of key differences among cells.

Analyses and Results: The unit of analysis in the results matrix is an MPP standard score for each replication in a cell. The transformation of a random number vector into one row of each of the 15 Z matrices is accomplished using an approach which comes close to maximizing the average correlation among the predicted performance scores for corresponding jobs across the cells of the results matrix. Thus the maximum power for a statistical test of the research questions can be obtained by using a repeated measures design in the initial, overall F test used to establish the significance of the differences among the means for which an overall test is appropriate. However, it is the tests of the critical differences between the cells, identified in advance, that profit most from the use of a repeated measures design.

The differences between the correlation coefficients for each index of the prescribed pairs of indices with a specified second variable (either PCE or PAE) will be computed as follows:

(1) H_d and PDI with PCE

(2) Modified H_d and modified PDI with PAE

(3) H_d and Max-PSE with PCE

(4) PDI and Max-PSE with PCE

(5) Modified H_d and Max-PSE with PAE

(6) Modified PDI and Max-PSE with PAE.

The same three cells of the results matrix will be used in computing both correlation coefficients whose difference is tested for significance. For example in (1) above, the cell in which H_d is maximized, the cell in which PDI is maximized and the special (neutral) cell will be used to compute both coefficients.

APPENDIX 5.B

CAPITALIZING ON THE MULTIDIMENSIONALITY OF THE JOINT PREDICTOR-CRITERION SPACE IN DEVELOPING OPTIMAL ASVAB COMPOSITES FOR JOB ASSIGNMENT AND COUNSELING

APPENDIX 5.B
CAPITALIZING ON THE MULTIDIMENSIONALITY OF THE JOINT PREDICTOR-CRITERION SPACE IN DEVELOPING OPTIMAL ASVAB COMPOSITES FOR JOB ASSIGNMENT AND COUNSELING

Overview: A factor analysis of the Project A experimental predictor pool extended into the criterion space is used to identify one PSE efficient set of factors and one PCE efficient set of factors. The classification efficient set includes k (either 3 or 4) PCE efficient factors (H_d maximized); the second set of k factors will maximize H_a. Both sets will span the joint predictor-criterion space defined by the weighted criterion components of 19 jobs (one for each of 19 sub-families). The two separate sets of k factors rotated to provide a simple structure with the eight jobs will be treated as if they were AAs.

The PCE of these two sets of carefully rotated k factors will be compared with the PCE provided by use of the existing 9 A~~ , aptitude areas with respect to these 9 job families. Each set of k factors will be used two ways: (1) to compute LSEs for each job, and (2) as factors rotated to a meaningful simple structure in the job space. Factor based composites will be separately derived from two kinds of rotated factors: one in which all 19 jobs are utilized to determine simple structure, and one in which simple structure is sought with respect to the existing job families.

The first of these two types of factor based composites will be combined into simply stated composites corresponding to each of the 19 jobs based on fewer factor scores than were used to compute the LSEs, and weighted by 1, 2, or 3; it is these combinations that will be used as assignment variables. The latter will be derived from each of the two factor solutions (one maximizing H_a and one maximizing H_d) by rotating the factors to match the existing major job families.

Thus there are three sets of composites derived from each of the two factor solutions (sets 3 through 7). The first set of composites consists of the 19 LSEs computed in the total space, and the eighth set consists of the Army aptitude areas. These eight composite sets, in the order they are listed in Table 5.B.2, follow:

1. nineteen LSEs based on total information;
2. nineteen LSEs based on k classification efficient factors;
3. nineteen LSEs based on k selection efficient factors;
4. k composites based on k classification efficient factors rotated to fit 19 jobs;
5. k composites based on k selection efficient factors rotated to fit 19 jobs;
6. k composites based on k classification efficient factors rotated to fit 9 job families;
7. k composites based on k classification efficient factors rotated to fit 9 job families;
8. nine Army aptitude areas;
9. nineteen composites based on k classification efficient factors; using weights of 0, 1, or 2;
10. nineteen composites based on k selection efficient factors; using weights of 0, 1, or 2;
11. nineteen composites based on k classification efficient factors; using weights of 0 and plus or minus integers 1-4;
12. nineteen composites based on k selection efficient factors; using weights of 0 and plus or minus integers 1-4.

PCE will be computed on a censored distribution resulting from the truncation of (selection on) AFQT. PCE will be computed for all eight sets of composites. Selection will be accomplished using the same SR (probably 0.70) and an LP program used to assign entities (vectors of synthetic scores) preliminary to the computation of PCE.

This model sampling experiment will utilize a cross validation design to assure an unbiased comparison of the existing Army AAs with the other seven sets of composites. PCE values will be computed for each entity sample. Utilities associated with the use of all eight alternative sets of composites will be computed and implications for using the factorially based composites for counseling in connection with high school recruiting will be considered.

Problem: Just as a set of tests can be selected to represent an experimental test pool, to maximize either H_a or H_d, so can a set (usually a smaller one) of factors or test composites be selected to represent such a pool of predictors in a specified joint predictor-

Table 5.B.2. Composites and Corresponding Transformation Matrices

Composite Numbers	Index Enhanced	Composite Identification	Transformation Matrix	Number of Columns in F_a or F_d
1	PUE	19 LSEs based on total information	F_a'	19
2	PCE	19 LSEs based on *k* classification efficient factors	F_d'	*k*
3	PSE	19 LSEs based on *k* selection efficient factors	F_a'	*k*
4	PCE	*k* composites based on *k* classification efficient factors rotated to fit 19 jobs	$T_{dr1}F_d'$	*k*
5	PSE	*k* composites based [illegible] selection efficient factors rotated to fit 19 jobs	$T_{ar2}F_a'$	*k*
6	PCE	*k* composites based on *k* classification efficient factors rotated to fit 9 job families	$T_{dr3}F_d'$	*k*
7	SSE	*k* composites based on *k* selection efficient factors rotated to fit 9 job families	$T_{ar4}F_a'$	*k*
8	PSE?	The 9 Army aptitude areas	$F_a'a$	19
9	PCE	19 composites based on *k* classification efficient factors; weights = 0, 1, or 2	$F_d'W_1$	*k*
10	PSE	19 composites based on *k* selection efficient factors; weights = 0, 1, or 2	$F_a'W_2$	*k*
11	PCE	19 composites based on *k* classification efficient factors; weights = 0 or signed integers 1 through 4	$F_d'W_3$	*k*
12	PSE	19 composites based on *k* selection efficient factors; weights = 0 or signed integers 1 through 4	$F_a'W_4$	*k*

criterion space. The issue of whether practical gains in PCE result when this selection of composites is specifically made so as to maximize PCE rather than predictive validity has two facets.

One facet is methodological: one school of thought contends that both PSE and PCE are best served by maximizing predictive validity; this position is disputed by others (including the authors) who contend that practical, as well as theoretical, gains in PCE are obtainable by selecting predictors to maximize PCE.

The second facet relates to the dimensionality of the joint predictor-criterion space. Many in the validity generalization movement believe that regardless of what psychometric theory might show regarding the advisability of attending to PCE in the predictor selection process, the joint predictor-criterion space is essentially unidimensional with at best 2 or 3 relatively trivial dimensions potentially available for addition to the general mental ability found in ASVAB. A major adherent to this movement contends that the Army aptitude areas are essentially unidimensional.

The Army AAs are needed as a baseline against which to compare the benefits and costs associated with the new sets of composites. The Army AAs have evolved from several research efforts since WWII and have been modified to reflect current data collected under the auspices of Project A (McLaughlin et al., 1984). In recent years an emphasis on predictive validity has dominated the considerations as to whether the content and number of AAs should be confirmed or modified. Unfortunately, the AA set which will be used as an operational base line has not been modified to reflect the new predictor development effort of Project A. Thus the present AAs are the best available, although not the best conceivable, set of operational composites for use as a basis of comparison in determining the utility obtainable from using alternative methodologies.

The use of a cross validation design avoids one type of bias (the presence of correlated sampling error in the assignment and evaluation variables), but does not avoid an era determined bias. The current set of AAs reflects a universe of an earlier era. Our currently available data with which we define universe values reflecting a current era may not predict the universe values of a future era any better than do the universe values proclaimed on the basis of large samples of data collected in an earlier era. We believe the data currently available to us are most excellent and certainly permit us to define the current universe with confidence; we still do not know the future universe. Parameters computed on cross samples drawn from a defined current universe have an advantage over parameters

computed on samples from a past "universe" since the evaluation process in our model sampling experiment of necessity uses parameters computed on an independent sample of the current "universe."

Research Questions: Research questions focus on the effectiveness of twelve kinds of test composites when used as assignment variables in an optimal personnel assignment algorithm. Each set of composites is a representation of the joint predictor-criterion space; effectiveness of a composite set in representing this space is determined by the magnitude of the MPP standard score after all entities are optimally assigned.

The purpose of this study, as defined by these questions, differs from that of the first study which uses tests, rather than composites, to represent the predictor-criterion space. Also, it is assumed in this study that an operational assignment system will make use of a much smaller number of test composites for recording in personnel records and for making both initial and subsequent career decisions, even if a larger number of LSEs are used to make initial assignments.

This second study differs from the third study (described below) in that the emphasis is on the predictor variables rather than on the assignment strategies; the research questions of the third study concern the effectiveness of alternative selection/assignment strategies in the context of hierarchical layering characteristics.

The research questions listed below will be more precisely expressed as hypotheses prior to actual collection of model sampling data; these research questions are:

1. Can a set of *k* classification efficient factors provide a larger amount of PCE than a set of selection efficient factors? Does the difference between the PCE obtained from: (1) LSEs computed in the space spanned by *k* classification efficient factors, and (2) LSEs computed in the space spanned by selection efficient factors, have statistical and practical significance?

2a. Which of two alternative approaches provides the best set of test composites (each composite, singly or in weighted pairs, associated with one of the 19 major job sub-families) for use in the classification of personnel? Does a set of 3 or 4 composites selected to maximize PCE for 19 Army jobs provide more PCE after assignment than can a set of 3 or 4 composites selected to maximize prediction effectiveness? Do one or both of these sets of factor based composites provide a statistically and practically significant gain in PCE over the use of the existing (as of 1988) Army aptitude areas (AAs)?

2b. Which of two alternative approaches provide the best set of test composites, as in 2a above, except that each factor based composite is associated with one of *k*

job families (each of the 19 jobs is placed in one of the k job families) for use in the classification of personnel? Do one or both of these sets of factor based composites provide a statistical and practical significant gain in PCE over the use of Army AAs?

3. Can a set of 3 or 4 test composites, used as singlets or as simply weighted pairs corresponding to each of 19 Army jobs, adequately approximate the classification efficiency that can be provided by 19 separate LSEs (one for each job)? Is the PCE provided by LSEs greater (with statistical significance)than that provided by a set of 3 or 4 optimally constructed composites used in accordance with simple rules?

4. Can a set of LSEs, one for each of 19 jobs, computed in the total space provide significantly more PCE than a set of LSEs computed in the more restricted space spanned by 3 or 4 PCE efficient factors?

5. Are there practical and statistically significant differences between the PCE values provided by the use of the eight composite sets that are consistent with the following hypothetized hierarchy of magnitudes: (1) LSEs computed in the total space > [(2) LSEs computed in the selection efficient space >?< (3) composites obtained in the classification efficient space and combined to relate to jobs] > (4) composites obtained in the selection efficient space and combined to relate to jobs > (5) composites obtained in the classification efficient space and matched to k groupings of the 19 jobs > (7) composites obtained in the selection efficient space and matched to k groupings of the 19 jobs > (8) Army aptitude area composites.

6. Does the set of 3 or 4 composites hypothesized to have the largest amount of PCE, when used in conjunction with the "g" composite, have the characteristics desirable in a set of composites to be used for counseling high school students who are considering a career in the Armed Forces? What is their reliability? Are they interpretable in terms of traditional factors commonly used to explain test content and the aptitudes of those receiving vocational counseling? What is the predictive validity and PCE provided by this candidate set of composites? How well does this composite set compare (in respect to the above considerations) with the test composites whose scores are currently offered to high school counselors?

Approach: This study divides into the following four stages: (1) the application of factor analysis techniques to intercorrelations and validity data from Project A to obtain parameter values used in the following steps; (2) the conduct of the model sampling experiment in which samples of synthetic score vectors (entities) are generated, a classification process simulated, and mean predicted performance scores computed for each

sample; (3) the analysis of the results from the model sampling experiment; (4) the conduct of a utility analysis and interpretation of results.

Obtaining Parameter Values

The validities of the nineteen Army jobs selected for use in validating the Project A experimental test pool will be corrected for restriction in range to provide for a youth population, the estimated validities of all the experimental and operational predictors against the criteria of the 19 jobs.

These corrected validities will be used in conjunction with the intercorrelations of the predictors (also corrected so as to represent a youth population)to compute a 19 by 19 covariance matrix among the predicted performance scores of the 19 jobs. This predicted performance covariance matrix is called C, and the corresponding correlation matrix called R_p, just as in Part 2 of this report.

Still in preparation for the main model sampling experiment, the matrix C will be factored to obtain the matrix F. Two samples of N by 19 random numbers (i.e., the matrix X) will be generated and two independent samples of predicted performance scores generated, each equal to XF'. Two separate covariance matrices, C, will then be computed; one C will be used to compute the parameters used in the predicted performance estimates used as evaluators and the other C for computing all other parameters discussed below.

The classification efficient factor solution, F_d, can be defined in terms of the roots and vectors of $(F-HF)'$ $(F-HF)$, where F is any matrix such that $FF' = C$, and H is defined as in Chapter 8. Since $T_o'(F - HF)'(F - HF)T_o = D_o$, $F_d = T_oD_o^{1/2}$. Only k columns of F_d will be retained (k will be set at 3, 4, or 5 depending on the number of factors that have at least two non-trivial coefficients. Thus F_d is a 19 by k matrix of factor coefficients.

The matrix F_d will be further rotated to achieve simple structure and F_{dr} defined as an orthogonal transformation of F; thus, $F_{dr} = FT_{dr}$, and $(T_{dr})'F'$ is the matrix that can be used to transform an N by 19 matrix of random numbers into an N by 19 matrix of least squares estimates of job performance.

The factor solution, F_a, both maximizes the contribution of successive factors in the joint predictor-criterion space, and provides a set of factor based composites which provide an approximate maximization of PSE. The matrix F_a is both the pc solution of C (i.e.,

$F_a F_a' = C$, and is the solution equal to $AD_a^{1/2}$ where $A'F'FA = D_a$, $AA' = A'A = I$, and D_a is the diagonal matrix of eigen values.

The first k columns of F_a are also rotated to simple structure and the orthogonal transformation matrix T, such that $F_a T_{ar} = F_{ar}$, retained. The same number of columns of F_d and F_a will be discarded prior to rotation.

Score vectors for the twelve sets of composites will be generated by multiplying a vector of random numbers $(x)_i$, by transformation matrices. Each of these transformation matrices is defined in Table 5.B.2.

The distinction between the generation of synthetic scores representing LSEs based on the first k factors of either F_d or F_a, and LSEs based on the total space is in the number of columns of F_d or F_a that are utilized in transforming the vector of random numbers.

The transformation matrix to be used in generating Army aptitude area scores is obtained by extending the factor solution, F_a, to the 9 aptitude area variables. In this 9 by 19 factor extension matrix (F_{aa}), the elements are the factor coefficients, the columns represent the pc factors as found in the joint predictor-criterion space, and the rows represent the Army aptitude areas. The transformation matrix is simply $(F_{aa})'$.

The Model Sampling Experiment

The model sampling experiment commences with the generation of a vector of random numbers which is then transformed into twelve separate vectors of predicted performance scores plus a synthetic AFQT score. Each of these twelve vectors is one entity in a sample of N entities. Approximately thirty percent of each sample will be rejected; all entities with an AFQT score below a specified cutting score will be deleted. The remaining entities in each such sample are then optimally assigned and the mean predicted performance standard score computed and placed in the results matrix as one replication within one of the twelve cells.

Each 1 by 19 vector of synthetic scores has an expected covariance matrix equal to the transformation matrix premultiplied by its transpose. As shown in Appendix 5.C, these entities have the same expectations as a real sample drawn from a universe with the indicated covariance matrix.

The MPP performance standard scores are based, not on the variable used as the objective function in the assignment process, but on the least squares estimate of performance computed from the total set of predictor variables (in the total space). This

LSE computed in the total space is used as the evaluation variable regardless of which assignment variable is utilized. LSEs used as evaluation variables are based on regression weights computed from an independent sample of entities as compared to the cross-sample used to compute the weights used to define the assignment variables.

Research Design: The MPP standard score constitutes the unit of analysis used to test the hypotheses derived from the research questions. This aspect of the approach will be discussed after the research design has been further considered.

Transformation matrices that provide the maximum correlation between the PCE values for the twelve entities generated from the same initial vector of random numbers will be used and a univariate analysis capitalizing on repeated measures used to test significance.

This essentially unifactor experiment has twelve levels with a hypothesized hierarchy of magnitude. However, only certain contrasts have relevance to the research question and the hypotheses to be statistically tested should relate only to these contrasts.

The research design, including the use of separate samples to compute the assignment and evaluation variables, is based on a particular model of reality. It is assumed that two independent samples drawn from the same universe as is represented by the Project A data would provide two sets of LSEs that differ to the same extent as two sets of LSEs computed from two samples generated from a designated universe defined in terms of the Project A data. We do not believe the amount of correlation error across the assignment and evaluation variables that would result from using the same large sample to compute the parameters of both the assignment and evaluation variables can seriously affect the results of the first and third of these three experiments. However, we feel that the comparison of a composite set that has no parameters derived during this study with other components whose parameters will be derived using the data of this study requires a cross-validation design.

The concept from which the cross validation design is derived makes the assumption that the empirical data provides a reasonable estimate of the universe intercorrelations among predictors and of the validity coefficients linking all predictors and jobs. It is also assumed that random samples consisting of synthetic entities drawn through model sampling techniques can represent the effects of sampling error on the inferences one would like to make about the utility resulting from the application of alternative classification strategies.

One set of model sampling entities, sample 1, can be considered to be a surrogate of the empirical sample from which it is hoped to make operational decisions. It is important that this sample have the exact same number of entities representing each job as there are individuals in the empirical sample. Sample 1 will be drawn (generated) so as to assure that the expected intercorrelation and validity coefficients equal the values obtained in the empirical sample. Sampling error will cause sample 1 to differ, on the average, from the empirical sample to the same extent that the empirical sample differs from the true, unknown universe.

A second sample, sample 2, can be generated so as to provide an independent estimate of universe parameters required to evaluate simulation results based on parameters computed on sample 1. The evaluation of the benefits resulting from each simulation will, of course, be in terms of an FLS estimate of performance using the regression weights computed in sample 2. Since logic calls for using the best available estimates of universe values in making these evaluations, a convincing argument can be made for making sample 2 larger than sample 1. This argument would logically lead to using actual universe estimates, rather than sample estimates, in the determination of the evaluation parameters. However, in order to simulate a cross validation design in which an available empirical sample is randomly divided into two equal halves to provide the equivalent of samples 1 and 2 as used in this study, sample 2 may be generated using the same number of entities for each job as is used for sample 1.

A third sample, actually a set of subsamples making up sample 3, can be generated to provide "cross sample" simulations. In these simulations all assignment variable weights are provided from sample 1, the synthetic scores to which those weights are applied are provided in sample 3, and the weights used to compute predicted performance as the evaluation measure are provided from sample 2. Each subset of this third sample is the sample of entities generated for each replication, under prescribed conditions, of the model sampling experiment. The comparisons of predetermined composites, such as the Army Aptitude Areas, with variables based on weights computed in sample 1, such as sets of factor scores, FLS composites, or LSEs based on selected sets of factor scores, will be unbiased in sample 3.

There is no need to compute intercorrelation and validity coefficients for sample 3. For each of these samples the PP scores of each entity for each job will be generated, each entity assigned to a job, and the weights from sample 2 used to estimate MPP scores for optimally assigned entities.

For implementation of results, the weights identifying assignment variables to be recommended for operational use will be recomputed using the empirical sample. This sample provides the best estimate of the universe values of these weights.

The objective function (i.e., the mean standard score for the assignment variable after assignment) will also be recorded for all samples in which the assignment variables are LSEs based on total predictor information. The extent to which the results would have been affected by using the objective function as a measure of PCE can be inferred from these data.

Analysis and Results: The hypotheses constructed to reflect the research questions will be tested for statistical significance. The hypothesis with the most practical significance is that there is no difference in the PCE provided by the composites based on F_d as compared to those based on F_a. The hypothesis with the next most practical relevance is that there is no difference between the PCE provided by the composites based on factors and that provided by the Army aptitude areas. The comparison of the PCE provided by the four sets of composites to that provided by the LSEs is of lesser practical importance, since the role of 19 LSEs in effecting initial assignments to specific jobs is quite different than the counseling type function that will always require a small number of test composites that have easily understood relationships to job families.

It is anticipated that the superiority of 19 LSEs over either 9 aptitude areas or a set of 4 test composites in the assigning of entities to 19 jobs will be readily established with a high level of statistical significance. However, the magnitude of the increased utility from using job specific LSEs in making these assignments is not known; the advantage provided by LSEs over aptitude areas reported elsewhere (Sorenson,1965) pertained to use of one LSE for a total job family.

APPENDIX 5.C

MULTIDIMENSIONAL SCREENING: COMPARISON OF AN ASVAB SINGLE STAGE SELECTION/CLASSIFICATION PROCESS WITH THE TRADITIONAL TWO STAGE PROCESS

APPENDIX 5.C
MULTIDIMENSIONAL SCREENING: COMPARISON OF AN ASVAB SINGLE STAGE SELECTION/CLASSIFICATION PROCESS WITH THE TRADITIONAL TWO STAGE PROCESS

Overview: A model sampling experiment utilizes parameter values obtained from a large empirical study (Project A) to determine utility outcomes from the application of a little known, although optimal, selection/classification strategy (MDS) to reject and assign synthetic individuals to a set of several jobs under carefully controlled situational variations, included as the impact of: (1) SR; (2) hierarchical selection and classification characteristics of a general predictor across several jobs; (3) hierarchical selection and classification characteristics of a set of job specific predictors (with the effect of "g" removed) across the same jobs; (4) hierarchical selection and classification characteristics using predictors that include a general predictor and a set of job specific predictors from which all hierarchical layering effects have been removed; and (5) hierarchical selection and classification characteristics of the LSEs across these same jobs. MDS is compared with: (1) traditional two stage selection and classification strategy in which selection is first accomplished on "*g*" and classification accomplished later on LSEs; and (2) the less traditional selection and assignment on "*g*" times the validity of *g* for the i^{th} job (acceptance accompanied by a quota constrained optimal assignment to a job). One of five situations to be evaluated includes a set of actual jobs, criteria, and predictors selected from the "Project A" study.

Problem: Most of the gain from combined selection/classification comes from the selection part when the percent rejected is non-trivial. Yet, traditionally, an abbreviated predictor is used for selection and the larger test battery reserved for a later classification effort in a second stage process, rather than seeking to make maximum use of all predictors to effect both selection and classification. Also, no operational use of the optimal, simultaneous selection and classification strategy (process) has been reported in the literature. Such an optimal process has been visualized in the depiction of selection classification models (Brogden, 1946b, 1959; Cardinet, 1959) but no empirical evaluation

of such a model has been described. An optimal algorithm, MDS, is described in Chapter 6.

While hierarchical layering effects on classification have been noted by several authors, very little is known concerning hierarchical layering effects on the selection process. Also, investigations of hierarchical classification effects have been sketchy with little attention given to the theoretical basis of the hierarchical layering effects and little effort made to attribute PCE to allocation and hierarchical classification separately.

It is easy to see the cost of using all available predictors in an optimal simultaneous selection/classification process, but utility cannot be computed until the benefits (gains) from using MDS is measured. The utility results from a particular empirical situation cannot be generalized to future situations or projected for planning purposes until benefits under different SRs and degrees and sources of hierarchical layering (with respect to either selection, classification, both or neither) are determined.

Research Questions: The following research questions pertain to Project A conditions and to additional conditions resulting from carefully controlled variations in model sampling parameters:

1. Are the gains from MDS statistically significant and large enough to offset additional administrative costs?
2. What is the effect of SR on utility?
3. What are the effects of hierarchical layering in selection and classification on the utility of using each of three alternative strategies? Which differences in effects are statistically significant and which parameter changes have a practical impact on utility?

Approach: The first step is to analyze Project A intercorrelations (R_t) and validity (V) matrices for a youth population to obtain the parameters for one of the five situations (described in terms of data characteristics) represented in the experiment. This step will be accomplished by obtaining the largest PC factor in the joint predictor-criterion space and calling this factor "g." Four or five additional factors in the residual joint predictor-criterion space, each of which successively maximizes H_d, will be rotated with the orthogonality constraint relaxed so as to provide the best simple structure for four or five (out of nine) jobs. The objective is to find k factors and k jobs for which a reasonably good oblique simple structure is obtainable (k = 4 or k = 5). These k rotated oblique factors will be called the u_i factors (the classification efficient factors corresponding to the job).

The second step is to compute the LSEs of each factor ("g" and the 4 or 5 u_i factors) and to compute the R_t and V matrices corresponding to these 5 or 6 predictor variables and 4 or 5 job criteria (for S1 through S5). The model sampling parameters for situation #5 (S5) will be derived from the initial pair of R_t and V matrices. The R_t and V matrices required to generate synthetic scores with the desired data characteristics for S1 through S4 will then be computed from factor extension matrices constructed to yield the same average intercorrelations and multiple correlations with the criteria as are present in S-5, but with systematic variations in hierarchical layering characteristics as described in the research design section.

Step 3 calls for generating synthetic score vectors for each entity (an artificial individual). Each score vector has an element (score) for each of the 5 or 6 predictor variables. These synthetic scores have expected intercorrelations and validities equal to the youth population values. Each entity will be immediately rejected or assigned by each of the three experimental strategies for two separate levels of SR; when a batch of entities has been assigned the MPP standard scores will be computed and six replication values provided for three cells in the matrix of results.

The next to last step (4) is the analysis of results. Statistical tests are computed on the matrix of results and MPP standard scores examined for trends and interaction effects. Statistically significant gains in MPP will be considered further in terms of dollar based utility afforded by each strategy.

The last step (5) is the computing of utility as a function of costs and benefits of each of the three strategies under each of the five data characteristics category. and for two levels of SR.

Research Design: Five separate values of R and V, each a k by k matrix, will be used to generate a N by k matrix of synthetic scores; from each vector of random numbers five separate entities (score vectors) will be generated. These five data characteristics are as follows:

S1. The "g" scores have hierarchical characteristics evident in the defining V matrix, but there is no hierarchical layering in the u_i.

S2. The "u_i" scores have hierarchical layering effects but there is no hierarchical layering present in the validities of "g."

S3. Hierarchical layering effects are present in both g and in the u_i (i.e., there is a variation in the validities of the LSEs across the k jobs with the source of this variation coming equally from g and the u_i.

S4. All g and u_i components of the job LSEs have the same means and variances, and thus none of the three strategies can capitalize on hierarchical layering.

S5. The g and u_i reflect the existing empirical relationship between selected job criteria and test composites.

The three selection and assignment processes are represented by three processes defined as follows:

P1. Selection on g, then optimal assignment on LSE.

P2. Simultaneous selection and assignment using g times r_{gi} as the selection and assignment variable.

P3. Simultaneous selection and classification using MDS and the LSEs.

Some thought should be given to the two levels of SR used to estimate the interaction of SR with the other variables in affecting MPP. The values of 0.75 and 0.50 or 0.80 and 0.40 are two credible alternatives.

The model sampling process proceeds from the initial generation of a vector of random numbers which is transformed into 5 score vectors (S1, S2, S3, S4, and S5). Six different processes are used either to reject or assign each of these entities (score vectors) to one of the k jobs. These processes are P1, P2, and P3 for two levels of SR. Thirty MPP standard scores result from the generation of a *single vector of random numbers*. If a sample size of 200 is used, the generation of ten sets of 200 random vectors would provide ten replications for each of the 30 cells in the "results" matrix. Two of the three processes (P1 and P3) require an LP solution of a k by 200 matrix for each replication, unless it is decided to use universe column constants and the "sequential" model for achieving an optimal assignment. It may be desirable to accomplish 5 replications per cell using LP assignments and then to use the resulting column constants to accomplish an additional 5 replications.

Analysis and Results: The unit of analysis in the "results" matrix is an MPP standard score for each replication in a cell. From an analysis of variance point of view the experimental design has 3 factors with 5, 3, and 2 levels respectively plus 10 replications per each of the 30 cells. F tests will be used to test the significance of both main effects and interaction effects. The replications will provide the error term. Of the 3 factors, only SR is based on a natural continuum.

All main effects can be expected to be significant using the proposed large number of replications, considering: (1) the theoretical superiority of a higher SR over a lower SR,

(2) the superiority of P3 over both P1 and P2, and (3) the theoretical superiority of P1 over P2 for all conditions except S1. P1 and P3 should theoretically have the largest MPP values for S2, then S3, then S5, and finally S1 and S4 as a possible tie. P2 would yield the following rank order of situations: S1, S3, S5, and finally S2 and S4 as a possible tie.

The primary purpose of the study is not to test the above preconceived relationships between the interactions of data characteristics and processes, but to provide estimates of the utility that can be expected from three alternative strategies. Thus once the "results" matrix has been established as unlikely to have occurred by chance, the important analysis is the computation of utility in terms of dollars.

Note that P2 is very close to the strategy frequently proposed by Schmidt and Hunter, and P1 is the strategy that the authors recommended as an intermediate improvement that would increase utility over the present approach assigning on composites (AAs) instead of on LSEs. The comparison of P3 with P1 and P2 is not tilting at windmills that no one would ever propose as dragons (candidate strategies). Thus P1 and P2 are appropriate strategies providing lesser benefits at lesser costs and are the appropriate competitors with which to compare MDS on the basis of utility.

APPENDIX 5.D

METHODS FOR RESTRUCTURING JOB FAMILIES

APPENDIX 5.D
METHODS FOR RESTRUCTURING JOB FAMILIES

Overview: Each of two major data sets available for analysis has important advantages for addressing several related problems whose solutions would add greatly to the knowledge of the psychometric principles of personnel classification. Project A data contains 19 scores for the 9 ASVAB tests plus 20 additional experimental tests administered to soldiers for whom performance criteria for 19 different jobs were also obtained. The 1981-1982 accessions make up what we will call the 81-82 data. These data include scores on the 9 ASVAB tests and 98 job criteria. The same 19 jobs of the Project A study are included among these 98 jobs.

The Project A data will be used to provide the covariances that, when corrected for selection effects, define a population of entities whose predictor and predicted performance (PP) scores yield the empirical covariance matrix corrected for selection effects. This corrected matrix estimates the covariance matrix of the youth population. This universe covariance matrix is used to generate two samples of entities with the same Ns for each job as found in the Project A data. One sample is the Project A analysis sample and the other is the Project A evaluation sample. The covariances of the 29 predictors bordered below by a 19 by 29 validity matrix, providing a 48 by 19 super matrix, is computed for each of these two samples.

Similarly, predictor and criterion scores for the 9 ASVAB tests and the same 19 jobs are extracted from the 81-82 data bank and a 28 by 9 covariance-validity matrix computed. This 81-82 analysis matrix will be used in the same way as the Project A analysis matrix, that is to (1) cluster jobs, (2) select jobs for use in the dimensionality tests, (3) compute composite weights for use in the cross samples for the creation of assignment variables, and (4) compute all other parameter values used in cross samples except the weights used for defining the evaluation PP composites.

The Project A evaluation sample will be used only as the source of the weights used to compute the evaluation PP scores. These scores, when aggregated for the entire cross sample, provide an unbiased estimate of MPP as a measure of potential classification efficiency resulting from the experimental conditions.

The cross samples will contain 216 entities each. These entities are generated using a transformation matrix derived from the Project A population covariance-validity universe matrix. The simulation of personnel classification using an LP algorithm is accomplished in each of these samples. Each entity is defined by a vector of PP scores. Entities are optimally assigned under quota constraints using assignment PP variables whose weights are defined from one or the other of the two analysis samples. These weights are applied to cross sample scores to create the assignment variables. Similarly, weights computed in the Project A evaluation sample are applied to the same cross sample scores to create the valuation PP scores that, when aggregated, provide the output (MPP scores) of each simulation.

Two methods will be used to cluster the 19 jobs into sets of 6, 9, and 12 families on the basis of two sets of data (1) the 81-82 sample, and (2) the Project A analysis sample. One clustering method will provide families that optimize PCE and the other will provide families that optimize predictive validity (PSE). A total of 30 replications, i.e., simulations using 316 entities each, will be used for each of the 24 experimental conditions.

The number of job families at each of 3 levels is determined to represent the current number of operational job families, 3 less, and 3 more. We would expect most of the aggregations of jobs into families to mirror the existing structure of sub-families within families. Although these 19 jobs may not provide a very sensitive test bed on which to determine the efficiency of the two alternative clustering approaches, the determination of the effects of reducing the number of job families from 19 to a smaller number (6, 9, or 12) should be based on the best available clustering technique. Hopefully, one of the two clustering methods being evaluated can be considered a close approximation to the "best" approach.

We predict that the results of this experiment will provide motivation for using the 81-82 data to confirm further the utility obtainable from using increased numbers of job families with associated FLS composites. Either simulations such as the one described in Chapter 3 or model sampling experiments using an LP algorithm to assign individuals must be included in such a confirmatory analysis.

The large number of jobs (89) possessed by the 81-82 data would qualify the 81-82 data bank as an excellent, comparatively sensitive test bed on which to make further comparisons of the two clustering methods. A further comparison of the two methods

using the larger set of jobs should be used. If a clear, statistically significant superiority of one method over the other is not evident in this experiment.

Problem: While the dimensionality issue is important to measurement theorists and has practical importance in the making of research decisions, the optimal number of job families is not a function of dimensionality. The utility obtainable from adding more job families is a function of whether the FLS composites can provide additional PSE and/or PCE in independent samples. A number of such composites could usefully exist in the context of a dimensionality of 1 by capitalizing on hierarchical layering effects, or with a dimensionality of 2 or more if no hierarchical layering effects are permitted.

Even so, we consider the dimensionality of the joint predictor-criterion space to be of direct interest to personnel system planners and to those responsible for identifying research opportunities. Also, we believe the credibility of personnel classification, and even of the use of test batteries by counselors, would be increased by a demonstration of dimensionality in the joint predictor-criterion space of at least 3.

The usefulness of job clustering is disputed in some circles because of the high correlation commonly found among PP composites and the instability of regression weights. We believe this is a result of defining the problem as one of whether jobs can be accurately categorized into clusters, rather than whether some jobs, or sets of closely related jobs, for which adequate validity data is available, should be treated as a separate family--as part of a strategy for increasing PCE. Whether the membership of large families can be accurately determined for borderline jobs is not really an important question, if there is sufficient validity data for these borderline jobs to justify computing an FLS composite for them alone.

If research results do not justify the use of clustering procedures based on empirical data, the conclusion should be that subsets of jobs with adequate validity data for computing FLS composites should not be combined into larger families. The more frequent conclusion is indicated--clustering into a few large families, and this clustering should continue to be based on expert judgment.

There is no reason to expect that nature has provided tightly clustered job families with sparsely inhabited border regions lying between families. It is just as likely that there are more jobs close to the boundaries between pairs of families than jobs close to the center of each cluster. Thus it is inevitable that membership in job families will change as data from independent samples are analyzed. We should not be distressed to find that jobs near

boundaries cannot be accurately placed into even large job families. Our concern should be with how each job can be accurately represented by an FLS composite to be used for initial personnel classification.

A practical question concerning criterion equivalence is whether decisions would be essentially the same if one criterion were to be substituted for the other. The Skill Qualification Tests (SQTs) and school grades available in the 81-82 data bank have been frequently criticized as inadequate for personnel research. There is reason to believe that the emphasis on discriminating between soldiers who almost, but not quite, achieved MOS or course standards, and those who just barely achieved these standards--has produced measures with very low ceilings, and for some jobs, little variance. The SQTs have the added disadvantage of having been constructed to diagnose training needs in addition to their evaluation function. On the other hand, the Ns are large for the 81-82 data and low criterion reliability is not necessarily a fatal flaw for use in making decisions when Ns are large. In this experiment the practical equivalence of results from the more economical 81-82 data with the Project A results will be examined in hopes that justification is provided for addressing operational classification problems requiring data on more jobs for a solution.

The increase in MPP available from the shredding of job families into more but smaller job families looks almost too good to be true. Confirmation using a cross validation design and carefully controlled independence between assignment and analysis variables in the cross samples is needed to obtain evidence that should be convincing to everyone.

Approach: The tests for dimensionality will be conducted much as described in the notional example provided in the text. The MOS samples to be utilized in this series of tests will be selected in each of the two analysis samples. Two MOS samples showing the largest differences in validities for their respective FLS composites (when computed separately in each sample) will be used to designate the MOS and provide the weights for the composites to be used in the cross sample analysis. Similarly, the best set of 3 MOS samples and, if needed, the best set of 4 samples will be identified in the two analysis samples.

The weights defining the FLS composites will be computed in each analysis sample and used to compute FLS composite scores in each cross sample generated for the model sampling experiment. A single large sample of entities, each defined by a vector of 9

synthetic tests scores and a synthetic criterion score, will be generated as cross validation samples to be used in conjunction with the 81-82 analysis sample. These 10 synthetic scores will have the same expected covariance matrix as found in the matrix defining the youth population.

Similarly, a separate set of cross samples will be generated for use with the parameters provided by the Project A analysis samples. The 29 weights computing using the Project A analysis sample results will be applied to vectors having 30 synthetic scores, representing the 29 predictors and the criterion score. Again, a separate cross sample of synthetic scores will be generated for each test of validity differences. The cross samples can, of course, be made as large as desired, but these Ns should, and will be, predetermined.

The generation of the 360 samples of 216 entities for determining the statistical significance of gains in MPP attributable to increasing the number of job families, using the better of two methods for clustering jobs, and for using the more efficient data source (one has more reliable criterion variables, the other more cases), will be accomplished much as in the previous three experiments. The research design for this aspect of the experiment is summarized in the text under the discussion of the "fourth experiment."

CHAPTER 6. RECOMMENDED CHANGES IN THE OPERATIONAL USE OF ASVAB

In this chapter we recommend changes in the way the ASVAB is used for personnel selection and classification along with the steps needed for implementing the changes.

Each year the military selects some 315,000 new recruits and decides in which job specialty each new recruit should be trained and assigned. Most of these recruits have little or no civilian work experience and consequently the services rely heavily on educational and aptitude test information.

The potential effectiveness of aptitude information, however, is greatly reduced in making job assignments for a number of technical and practical reasons. Among the technical reasons are: (1) the imposing of policy constraints, e.g., quality goals, that limit optimizations of predicted performance in allocating personnel to jobs; (2) the using of aptitude composites with poor differential validity for the classification process; (3) the employment of low minimum cutting scores in making assignments rather than using ordered lists of recruits based on predicted performance in meeting job quotas; and (4) the emphasizing of operational simplicity in assignment procedures, rather than the use of efficient computer-based algorithms for matching personnel to jobs.

Assuming that technical inadequacies in ASVAB composites and in algorithms used to make recommendations for assignment were to be resolved, this alone would not assure that the full benefits of an optimal job assignment system would be achieved--unless predicted performance information were utilized as the objective function in assignment algorithms utilized in actual practice. Therefore, in making recommendations for changing the assignment system to maximize mean predicted performance under the constraint of meeting quotas, we assume the acceptance of that goal by policymakers and also assume that most recruits can be persuaded to accept those jobs that they can perform best or nearly best.

On the basis of our analysis, we conclude that very large productivity gains can be achieved principally by changing the policies and procedures that govern the operational selection and assignment system. The initial changes we propose call only for the best use

of information in the present ASVAB. Later changes call for the addition of new job families for classification purposes and for the possible addition of classification-efficient tests in a revised ASVAB.

We propose a sequence of changes that are implementable over a period of several years, provided our assumptions and estimates are confirmed in the specific decision-context of each service.

Enhancing potential classification efficiency (PCE) processes results in large performance gains in the use of the ASVAB. Procedures that increase either allocation efficiency alone or hierarchical layering, through the use of an optimal assignment algorithm, increase mean predicted performance on the job. As noted earlier, the allocation process capitalizes on differential validity; hierarchical layering capitalizes on heterogeneous validities and/or job values that are reflected in the predictor variables used in the assignment process. The current Army aptitude area (AA) composites do not, however, elicit the hierarchical layering effects because the AA composites were standardized to have equal means and variances and are not weighted by either validity or job values. The recommendations we propose in this chapter for operational changes in the use of ASVAB information grow out of the application of sound psychometric principles to the estimation of utility gains obtainable from improvements in allocation efficiency and/or hierarchical layering. A number of changes we propose for implementation in the near future are simulated in Chapter 3. The desirability of these changes is confirmed by the utility values obtained from our realistic simulation. Other changes are proposed in Chapter 5. For the most part they are based on the results of prior studies. Others that were derived from psychometric principles require further confirmation from studies in progress or to be initiated shortly.

While we are confident that proposed changes based on prior results or psychometric principles that anticipate findings of studies in progress will confirm important additional gains, more precise estimates of gains and specifications of procedures are essential, apart from confirmation, before actual implementation can be initiated. Verification of nearly all the proposed changes should be available within the next year.

The sections below propose changes in the use of ASVAB to improve:

(1) Allocation efficiency

(2) Capitalization on hierarchical layering effects

(3) Specification of minimum job standards

(4) Measurement of the accomplishment of quality goals

(5) Selection efficiency

(6) Job family clustering (first to provide more families and associated composites, later to provide job clusters with more PCE)

(7) Recruit counseling and career guidance (implementing assignment recommendations based on maximum predicted performance)

(8) Algorithms for person by person assignment to maximize MPP

(9) Optimization of an integrated selection/classification process

A. IMPROVEMENTS IN ALLOCATION EFFICIENCY OF THE ASVAB

1. Use FLS Composites in Standard Score Form

The use of full least squares (FLS) predictor composites provides the maximum amount of PCE in a fixed battery. If FLS composites in Army standard score form were used, the capability to capitalize on hierarchical layering effects would be removed. Such composites would be comparable to the existing aptitude area (AA) composites that also have equal means and variances. However they would provide an assured increase in allocation efficiency over the present AA composites. We estimate, on the basis of prior studies and our simulation results, that such FLS composites may provide as much as a 50 percent increase over the present AA composites.

We suggest the use of FLS composites in Army standard score form because their use could be effectuated immediately (but only as a transitional measure); no policy changes would be required and the change to FLS AA composites would be transparent to operational personnel once computed, and thereafter would remain invisible to all.

We believe that the current relatively ineffective unit-weighted, three-test AA composites were initially adopted because of simplicity in their computation and use in a pre-computer age and possibly because researchers were not aware of the full impact of FLS on allocation efficiency. To transform FLS composite scores into AA scores, only a few computational steps are required. These steps are: (1) convert FLS composite scores into standard scores with a mean of zero and an SD of 1; (2) divide by the multiple R for each composite; (3) multiply by 20; and (4) add 100. These steps result in composite scores with means of 100 and standard deviations of 20, as is the case with the current Army AA composites. The AA composite names are retained. Weights for the FLS equations could be obtained from the simulation study, Table 3.10, but a more precise

estimate can be obtained from use of all project A data combined with the other information used to compute the weights for the simulation. As appropriate, weights could be adjusted by ridge analysis and/or other similar techniques. We recommend the use of weights provided by each of the services.

2. Use FLS Composites Converted to Predicted Performance

As noted at the start of the previous section, FLS predictor composites provide the maximum amount of PCE in a fixed battery. To take full advantage of both allocation efficiency and hierarchical layering, the FLS composite AA scores in Army standard score form are converted to standard scores with a mean of 0 and a SD of 1; these standard score composites are then multiplied by the multiple correlation coefficient (R). In addition to these FLS composites equal to predicted performance, as used in the assignment process, FLS composites in Army standard score form would continue to be used for records in the visible system and for all personnel decisions made after initial assignment.

On the basis of our simulation study, our conservative estimate of the performance gain provided by the FLS composites equal to predicted performance over the current AA composites is 73 percent. Prior study results (e.g., Sorenson, 1965), and the conservative procedures and estimates used in the simulation, lead us to consider a better estimate of the gain to be about 100 percent.

The use of predicted performance FLS composites require a change in policy, as do all of our remaining proposals. We consider such a policy change to have technical merit essentially since our simulation results show the use of FLS composite scores equal to predicted performance would only minimally affect the capability of achieving the quality distribution of individuals across jobs as prescribed by existing policy.

3. Use Classification-Efficient Tests

The change to FLS composites immediately provides the maximum available PCE in the present ASVAB and job families. Further improvements can be accomplished by the selection of new tests high in differential validity through the use of indices that measure PCE, to comprise an operational battery with the best available PCE.

An ongoing research effort is aimed at determining the MPP gain that may be achieved by sequentially selecting ASVAB and experimental tests validated in Project A in order to maximize PAE in a revised operational battery. We expect, on the basis of prior

findings (Harris, 1967), to find at least a 15 percent gain in MPP in use of new FLS composites comprised of classification-efficient tests over the present FLS composites. We expect a 10 percent gain from using an PCE-efficient index to select tests over the selection of tests to maximize predictive validity.

Implementing changes in the tests of the ASVAB requires a policy change that affects all services. Providing utility estimates obtained in the ongoing research effort should facilitate the decision process.

B. IMPROVEMENTS IN HIERARCHICAL CLASSIFICATION EFFICIENCY OF THE ASVAB

1. Use Predicted Performance Composites

FLS composites equal to predicted performance provide a maximum capitalization on hierarchical layering by reflecting the varying validities of the composites to achieve hierarchical classification efficiency. The existing AA composites cannot capitalize on hierarchical layering effects and must consequently provide a smaller MPP score. Also, the mirroring of validity effects in both the FLS assignment variables and the evaluation variables are guaranteed to increase MPP (with a small bias in our simulation due to correlated error across the assignment and evaluation variables). However, we keep the overall effect of the estimates low by conservatively estimating validity vectors and by using AA scores instead of test scores in the simulation. It was for this reason our simulation results showed the conservative estimate of 73 percent gain through the substitution of FLS composites that are equal to predicted performance and have disparate means and variances across jobs--as compared to the existing Army AA composites.

2. Use Job Values in Weighting Composites

Assuming that policymakers are willing to assign importance or values across different jobs and/or values for different performance levels in a job (Nord and White, 1988), such value weights could be used to convert MPP scores to new benefit scores. The use of benefit scores in the FLS assignment variables and in the evaluation variable increase hierarchical layering effects and therefore should increase MPP by at least 15 percent.

Research at ARI on job values is under way and should result in specification of how weights can be determined and used. However, it is unclear whether or not major

policy changes required to use these weights would be forthcoming. Thus we look for incorporating job value weights in FLS composites as a possible change in the long term.

C. RAISE MINIMUM JOB STANDARDS CUTTING SCORES

If ordered lists of recommended assignments for recruits based on predicted performance were actually used in the operational system (with the goal of approximating the optimization of performance as the objective function of an optimal assignment procedure while meeting quotas), minimum cutting scores for each MOS could be retained as the "basement" scores below which the higher cutting scores used to form ordered lists could not fall. But in actuality ordered lists are not used, rather recruit preferences and low minimum job standards are used in making assignments.

Our simulation showed that current job standard minimum cutting scores for job assignment should be raised by at least an average of five standard score units, resulting in a productivity of about 21 percent over current standards. However, this gain is only a transitional gain that reflects poor optimization of predicted performance in the actual operational assignment system being used at present. A future optimal assignment system based on predicted performance (PP), as described below, is expected to greatly diminish the role of minimum cutting scores in setting job standards. The use of cutting scores can only reduce MPP in an adequate assignment procedure. However, until the future optimal assignment system is implemented, raising minimum cutting scores is an effective and simple means of achieving productivity gains.

D. USE OF FLS COMPOSITES AS MEASURES OF RECRUIT QUALITY AND IN SETTING STANDARDS

1. Use FLS Quality Goal Measures in Assignment

The present assignment system uses a set of AFQT-based quality goals to provide a minimum percentage of AFQT category I - III A accessions in each MOS. AFQT, a measure of ability, is not, however, the best measure to use for this purpose. The FLS composites, each the measure of predicted performance for a job family, can be used for this purpose in place of AFQT. Quality constraints act to reduce optimizations, but quality constraints based on FLS composites, as compared to the use of AFQT, should increase the predicted performance for any job family in which the effect of the constraint is to increase "quality," and should decrease the competition for quality among families.

The superiority of the FLS AA composites, over of a single general composite also used for selection is more pronounced when the selection measure has maximum selection efficiency (maximum PSE)--as a general FLS or "g" composite would have. The FLS job family-specific composites would necessarily have a lower average correlation with each other than with a FLS "g" composite, thus reducing the competition for quality when quality is measured by the FLS job family-specific composites instead of by the FLS "g" composite.

However, a very inefficient selection composite that correlates poorly with both the FLS composites and the job criteria would provide less effect on the MPP scores of *other* jobs when the quality input must be increased to meet quality goals for a job. Thus, while the inefficient quality measure provides a smaller increase to the MPP of the job to which quality was input, as compared to the efficient measure, it has less effect on the MPP scores of all other jobs. In this way AFQT has an advantage over the use of the FLS job family-specific composites for some quality control strategies; the use of random numbers would have an even greater advantage in the same way and for the same reason.

The strategies used for meeting the quality distribution goals make a difference in the effect various measures of personnel quality have on the reduction of the objective function. One strategy relies entirely on restricting the supply of quality personnel into jobs that have no quality problems. The supply for jobs that have quality problems is thus improved and, hopefully, the quality distribution would be increased sufficiently to meet quality goals. A second strategy calls for actively channeling more high quality personnel to these jobs, possibly by resolving ties and near ties of adjusted scores for higher quality personnel in favor of the jobs needing more higher quality personnel.

Intuitively, under the first strategy, the use of a variable for effecting quality control that correlates poorly with PP scores will place less of a constraint on the objective function (the MMP score). In fact, reliance on this strategy could result in less of a reduction of the objective function--at the cost of making minimal changes in the MPP scores for the jobs which had their quality distribution "improved." It is likely that the smaller reduction in MPP we would expect from substituting FLS composites for AFQT in applying the second strategy would disappear or be reversed, if instead, the first strategy were to be utilized.

At present, we are unable to estimate the percentage of gain in classification efficiency obtainable from the use of FLS composites as compared to the use of AFQT to effect the required constraints. A simulation is needed to make such an estimate.

However, the logic of defining the term "quality" consistently and more precisely as it pertains to job assignment is alone sufficient to propose the use of quality goals based on FLS composites for early implementation.

This change could be implemented as soon as it is feasible.

2. Use FLS Quality Goals for Forecasting Personnel Quality Requirements in Future Systems

Quality goals based on FLS composites could profitably replace AFQT-based quality goals for specifying personnel quality requirements in future systems for the same reasons stated above. Additionally, use of such FLS composites, compared to AFQT composites, may facilitate the placement of new or modified jobs of a new system in the most appropriate job family, improving PCE.

However, as more job families are added and the two-tiered system (described below) is installed, the "visible" factor score composites, rather than the job sub-family-specific FLS composites, should be used to state quality goals. In the meanwhile it would seem practical first to substitute a FLS "g" component for AFQT. Such use would be transparent to the user, once readily computed statistical information on the youth population for the FLS "g" composite is provided to the materiel development community. Thus for the purpose of simplicity in using FLS-based quality goals for new systems, we proposed the use of a FLS "general" composite score based on a composite with weights that maximize the average validity for all jobs; the validity of this measure for each job is weighted by the number of operational accessions in each job to obtain the value that is maximized by these weights.

The visible scores used by personnel system users and the examinees themselves, however, are the FLS composite scores that have been converted to Army standard scores, making them appear to be identical to aptitude area composites.

Implementation of this change requires weights for use in the generalized FLS equation to form the FLS "g" composite. These weights can be approximated from the results of an ongoing research effort. It is necessary for each service to move systematically by computing weights and providing standardization data prior to implementation.

3. Use FLS Composites for Specifying Minimum Job Standards Cutting Scores

We have proposed the retention of "basement" or minimum cutting scores in making job assignments. FLS composites equal to predicted performance should be used in place of the present aptitude area composites in expressing these minimum standards.

E. USE THE GENERALIZED FLS COMPOSITE TO MAXIMIZE POTENTIAL SELECTION EFFICIENCY

Most of the gain in MPP from a combined selection/classification system comes from selection when the operational selection ratio is close to 80 percent, as it is presently. Yet traditionally, an abbreviated predictor composite is used for selection while the entire test battery is reserved for a later classification effort in a second stage process. To make maximum use of the battery, all predictors should be used in test composites for both selection and classification. (We refer to such measures as FLS composites when they are also LSEs.)

The use of an FLS general composite for selection and the use of a differently weighted FLS composite for each job maximizes the potential efficiency of both selection and classification (PSE and PCE) in a fixed battery.

We propose the use of a FLS general composite score, described above, to maximize predictive validity (i.e., PSE) in selection. Although a substantial gain over the presently used AFQT composite is virtually assured, a simulation study is needed to confirm that the difference in effectiveness between the two composites has practical significance in terms of utility. Such a study is under way, using the validity information obtained in Project A.

F. USE ADDITIONAL AND RESTRUCTURED JOB FAMILIES

A worthwhile improvement in PCE can be obtained by a major increase in the number of efficiently determined job families. An increase in the number of predictor composites and associated job families to somewhere between 20 to 40 would most likely provide the maximum efficiency for Army jobs--assuming data are available for computing moderately stable FLS weights for each family. Employing Brogden's (1951) formulations, we estimate that the performance gain resulting from such additional job families with their associated composites may be around 50 percent for an initial increase of families from 9 to 15.

A proposed research effort to be initiated shortly will employ optimal clustering algorithms (maximizing PCE) to shred the existing job families into a greater number of families with a smaller number of MOS in each job family. We recommend increasing the number of job families in phases. Initially, research will use the Project A data bank of 19 jobs to simulate shredding the existing job families into selected sub-families that have their own test composites. A subsequent step will use a training data bank of 98 Army jobs (McLaughlin et al., 1984) and a synthetic validity bank (Wise et al., 1988) to further shred out job families. A final effort using all available information will reconstruct the FLS composites and their associated job families.

Because of the large number of composites, the reconstructed system would be too cumbersome for all operational uses presently made of AA scores placed in a soldier's official file. We would defer implementation of an increase of test composites beyond 12, assuming research confirmation, until a two-tiered system is developed that permits concurrent use of the enlarged initial assignment system with the use of a smaller number of factor based AA scores.

G. USE OF A TWO-TIERED SYSTEM

One approach to utilizing 20 to 40 FLS assignment composites is to establish a separate system using about five or six factor score composites (FLS estimates of classification-efficient factors) to comprise the visible portion of the operational system. The first tier of the system would use the actual FLS job family-specific composites in a computer-based system to make assignment recommendations. This tier is transparent to the counselors and invisible to the recruit. The second tier of the system would enter factor score composites in the official records of each recruit in place of the AA scores now used. These factor scores can be used for recruit counseling, setting minimum cutting scores for entry into special training, and for other personnel management practices, such as career planning. We readily concede that detailed impact analyses of this proposal must precede a commitment to implementation.

Assuming the dimensionality of the joint predictor-criterion space is no more than 4 or 5, it is possible to define 1 selection-efficient and 4 or 5 classification-efficient FLS factor score composites. A study is under way to define these factors and to compare the PCE obtainable from optimal FLS composites, FLS Army standard score composites, factor scores, and FLS composites of factor scores.

Implementation of the two-tiered system may be defended on the basis of the amount of PCE retained in using a small number of factor composites simply combined to effect assignments (a simulation of the counseling process). We are confident of the technical feasibility and practicality of using factor composites, but it should be remembered that implementation of the enlarged assignment system described above depends on the use of a two-tiered operational system.

H. USE IMPROVED PERSON-JOB MATCHING ALGORITHMS

In Chapter 1, we described optimization procedures that maximize the mean assignment variable score and minimize the discrepancy between trial quotas and desired quotas. Most optimal assignment procedures operate in the context of a fairly complex set of constraints based on policy and practical considerations that reduce their efficiency. In the section below we make suggestions to improve the optimality of assignment procedures.

1. Use Predicted Performance and Attrition as the Objective Function

The services' operational assignment systems use aptitude area scores as assignment variables, in theory or practice. One consequence of our simulation is that the researchers involved in the development of EPAS are prepared to use predicted performance FLS composites as the EPAS assignment variables instead of the existing AA composites, upon approval of policymakers. The degree of impact that recommended assignments based on PP has on assignments actually accepted by the recruits remains to be determined.

If FLS composites are successfully substituted into EPAS, a first step in the installation of a two-tiered system will have been accomplished. The existing AA composites will continue to be used in the second tier until a smaller number of FLS factor scores can be utilized operationally.

As noted in Chapter 4, research findings indicate that retention can be effectively predicted by a composite comprised of such variables as aptitude, education, age and gender. Further, it may be possible to reduce attrition significantly while retaining most of the gains in predicted performance. Thus it appears desirable to utilize an objective function that maximizes predicted performance and minimizes attrition. A simulation study is necessary to determine the extent of assignment benefits before consideration is given to implementation.

The key to improvement of military assignment systems is to convince policymakers that even what appear to be small gains in MPP can translate into hundreds of millions of dollars in increased productivity each year. To realize such net benefits, policymakers must commit the services to the use of an optimal assignment system that maximizes performance (and hopefully minimizes attrition) as the objective combined function, while meeting job quotas, quality goals, costs and other constraints.

2. Use Person-By-Person Assignment Procedures

The advantage gained from the use of an optimal assignment procedure is reduced as the batches are reduced in size. An alternative is to simulate a large sample of artificial individuals (entities) defined by synthetic scores that have the statistical characteristics of the actual or projected input. Then, using the known requirements for each MOS, compute the dual solution parameters, i.e., column constants. These estimated column constants can then be used, one person (applicant or recruit) at a time, to identify assignments that maximize MPP in the defined population.

This technique can readily be incorporated in an operational person-job matching system such as EPAS by frequently updating the allocation plan, e.g., once every two weeks, and by the addition of column constants to each recruit's FLS score for each job.

3. Use Flexible Cutting Scores in Making Assignments

A matrix of adjusted assignment variable scores may be visualized in which each row of the matrix corresponds to a person to be assigned; each column contains the scores of each person for an assignment composite adjusted by the appropriate additive constant associated with the given job family. Then each individual can be assigned to the job corresponding to his/her highest adjusted score to maximize MPP. Such a set of assignments accomplished one at a time will, over an interval of time, closely approximate all quotas and maximize mean predicted performances. A primal solution for the simulated input sample can be readily made to provide the dual solution parameters, i.e., the column constants.

Cutting scores related to the column constants could assist the counselor to make recommendations to the recruit that are beneficial to both the recruit and the Army. In the counseling process, a column constant corresponding to each job family is added to the recruit's AA score to provide an adjusted AA score. This computation is performed each time a set of adjusted scores for a recruit are provided to the counselor. If each recruit were

to be assigned to his highest adjusted score, the total set of personnel assignments would be optimal, the MPP would be maximized. The required column constant should be periodically recomputed to reflect changes in expected input and requirements.

An interim assignment system that relies entirely on optimal use of cutting scores to effect MPP gains should be considered. Cutting scores for such a system would be set for adjusted scores, computed as above. Flexibility for considering recruit preferences could be provided by lowering the cutting scores in proportional amounts implied by the column constants.

When there are disparate selection ratios across jobs, we would expect to find in an all-volunteer force that the jobs with the best selection ratios have the best quality of applicants, and have the highest percentage of applicants who refuse to accept alternative assignments. For such jobs it does not make sense to use a first-come-first-selected policy in conjunction with very low minimum cutting scores. The cutting score for such jobs should be raised to permit the enlistment of the most qualified applicants, even if they are not the first to apply. Flexible cutting scores as described above can be used in conjunction with the existing minimum standards expressed in terms of AA composite scores.

I. USE AN INTEGRATED MULTIDIMENSIONAL SCREENING (MDS) SYSTEM

Appropriate column constants representing an applicant population can be applied to FLS composite scores to make both selection and assignment decisions simultaneously. Rather than selecting applicants using a single composite to provide a pool of recruits who are then assigned to jobs through use of FLS composites in a distinct second stage, the applicants can be simultaneously considered for acceptance and use in each job family. It can be assured through the use of the MDS approach that no person in the rejected group has a higher predicted performance score than anyone selected and assigned to any job family.

To understand the MDS algorithm, first visualize a matrix of assignment variable scores in which the rows represent the applicants and the columns the jobs with quotas greater than zero. In the MDS process an appropriate constant is added to each score in a column. The sum of the assignment variable score in each matrix cell and the column constant is the adjusted score. The largest adjusted score in each row of the score array is retained; the remaining scores are deleted. The retained scores are then visualized as placed in sort within each column and a cutting score set to accept just enough people to meet job

quotas. The selection-classification decision process can be made, one person at a time, using these cutting scores for each applicant. In addition to providing an optimal assignment every bit as good as obtainable from any LP solution, the counselor or computer making the selection-classification decision is provided the rank order of each job family for each person in terms of the contribution each assignment would make towards maximizing MPP, the objective function.

Prior research findings and psychometric principles both indicate that the use of MDS will provide dollar gains of practical magnitude, but the estimation of gains must be more precisely measured and systems features specified before the implementation of such a major policy change is recommended. A model sampling experiment has been initiated that will provide the needed estimates.

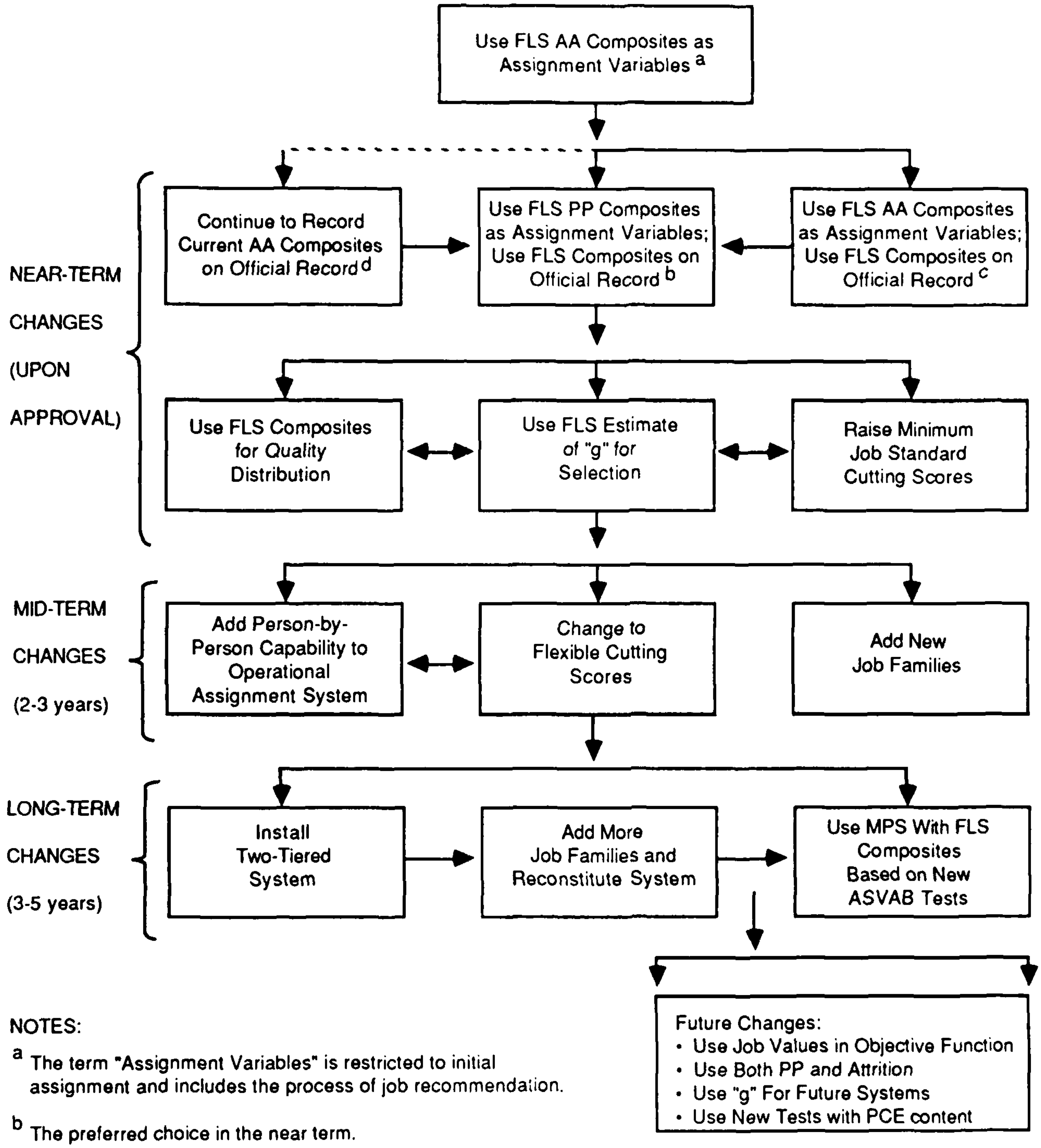
J. SEQUENCE FOR IMPLEMENTING OPERATIONAL CHANGES

In the preceding section we proposed a series of changes in each of nine areas. We indicate that the implementation of some changes can be made immediately, that some other changes require system development and testing before implementation, and that still others require additional research information to obtain estimates of gain and more precise specification of parameters.

We propose a number of technological improvements in the operational selection, classification and allocation system beyond those changes confirmed by the simulation results. All the recommendations we make, however, could almost certainly provide immediate benefits if implemented today, using available parameter values and procedures, but probably should not be installed until further research and management analysis are accomplished on how to make the most efficient applications of the proposed new procedures.

Figure 6.1 shows the sequence of change over three time periods. Some changes could be implemented in the near term after management analysis and policy approval. The sequences shows some other changes to be implemented within a two to three year period to allow for more precise estimates of productivity gains and more detailed specifications of how the changes are to be operationally investigated.

Still other changes require a three to five year period for implementation because they are dependent on the adoption of previous changes and completion of research and/or management analyses.



NOTES:

[a] The term "Assignment Variables" is restricted to initial assignment and includes the process of job recommendation.

[b] The preferred choice in the near term.

[c] The transition choice, before adopting the preferred choice.

[d] A temporary measure used for convenience during period of change.

Figure. 6.1. Sequence of Changes in the Proposed Selection-Classification System

In the near-term, the figure shows the preferred choice of using FLS composites based on PP as the assignment variable and as being recorded in Army standard score form for the official record. Because this is a major change in procedures and policy, and would require a management analysis of impact, we show an alternative as a transitional choice. The figure shows the use of FLS AA composites as the transitional choice for assignment and as being converted to FLS AA composites. The transitional change would not require policy approval but would require conversion to Army standard scores. The conversion would be transparent to operational personnel and invisible to others. Because this alternative still requires a conversion of scores, the figure also shows (with a dashed line) the less desirable measure of using FLS AA composites for assignment, but retains the current AA composites for recordkeeping. We regard this alternative only as a temporary convenience while preparing for the operational use of preferred FLS PP composites.

K. ESTIMATING THE GAIN FOR A CLASSIFICATION-EFFICIENT ASVAB

It is difficult to estimate accurately the aggregate present net value accruing from the adoption of our proposals. Any estimate made at this time is obviously a ball park figure. Our ball park estimate of gains attributable to improved operational procedures (to increase PCE) exceeds 200 percent in the aggregate. The largest contributor to PCE gains are FLS predictor composites, next are enlarged and restructured job families and then the addition of classification-efficient tests to the battery.

A few of the recommended changes are not additive gains. For example, gains obtainable from improved use of cutting scores are eliminated by the implementation of an optimal assignment algorithm combined with effective persuasion of recruits to accept their "best" assignments. Similarly the gain from substituting a FLS "g" composite for AFQT in selection is no longer relevant when an MDS system is implemented; the relevant gain becomes a combined selection-classification gain. The benefits for those changes require an earlier change that may not be additive to the earlier benefits. Most notably, estimates of gains obtainable from adding more job families are based on the assumption that both the existing nine-family and the fifteenth-family assignment procedures use FLS composites. It is obvious that successive improvements in job structure or test battery content are not additive; one gain is substituted for another as changes are completed.

The precise amount of dollar savings is not as important as are the relative differences in mean predicted performance among alternative strategies. We know from

our simulation results that improvements of one or two tenths of a standard deviation of MPP may result in very large gains. For example, a 0.143 gain in MPP provides more than a $260 million gain each year by substituting FLS composites for the current AA composites.

Although our simulation was accomplished with Army data, and our analysis of other data focused on the Army context, we feel the proposed changes are equally applicable to all services. We expect comparable gains, but suggest confirmatory analyses be conducted by each service.

Our analysis shows that the current Army AA composites are of limited value, but we also show that considerable classification efficiency is potentially obtainable from the present ASVAB if the battery is used in accordance with classification-efficient procedures. We believe the ASVAB would possess even more PCE if its development had not been largely based on a search for increasing the predictive validity of aptitude tests rather than on procedures for increasing MPP. Thus, we are not pessimistic regarding the future of tests developed for use in classification batteries. We acknowledge that PCE is difficult to achieve unless specific efforts are directed at developing predictors, identifying efficient tests for the battery and designing procedures that have, as their goal, increasing PCE. The proposed changes we suggest are directed principally at operational system design features that offer almost certain promise of large improvements in selection and classification efficiency.

GLOSSARY

ability test[a]--A test that measures the current performance or estimates future performance of a person in some defined domain of cognitive, psychomotor, or physical functioning.

achievement test[a]--A test that measures the extent to which a person commands a certain body of information or possesses a certain skill, usually in a field where training or instruction has been received.

adaptive testing[a]--A sequential form of testing in which successive items in the test are chosen based on the responses to previous items.

algebraic variability derivation--A technique for incorporating uncertainty into utility by the use of variance estimates.

allocation efficiency--The gain in benefit over random assignment obtained from an optimal assignment process attributable to differential validity.

allocation process--Classification that capitalizes on differential job validity.

alternative[c]--A course of action whose selection may result in an outcome that will attain the original objective.

aptitude test[a]--A test that estimates future performance on other tasks not necessarily having evident similarity to the test tasks. Aptitude tests are often aimed at indicating an individual's readiness to learn or to develop proficiency in some particular area if education or training is provided. Aptitude tests sometimes do not differ in form or substance from achievement tests, but may differ in use and interpretation.

assessment procedure[a]--Any method used to measure characteristics of people, programs, or objects.

attenuation[a]--The reduction of a correlation or regression coefficient from its theoretical true value due to the imperfect reliability of one or both measures entering into the relationship.

battery[a]--A set of tests standardized on the same population, so that norm-referenced scores on the several tests can be compared or used in combination for decision making.

behavior[b]--Observable aspects of a person's activities.

benefit--A theoretically desirable measure of performance that is value-weighted for jobs and validity in terms of an appropriate metric; when the benefit measure is correctly combined with costs, it provides a measure of utility.

break-even values--The determination of the lowest value of any individual parameter that would still yield a positive total utility value.

classification--The matching of individuals and jobs in an organization with the goal of maximizing aggregate performance; it requires multiple predictors jointly measuring more than one dimension and multidimensional job criteria.

classification[a]--The act of determining which of several possible job assignments a person is to receive.

classification battery-- A battery of tests used operationally to classify personnel.

classification efficiency--The gain in benefits over random assignment obtained from an optimal assignment process attributable to allocation and hierarchical classification efficiency; a separate LSE must be used for each criterion.

cognition[c]--The act or process of knowing, including both awareness and judgment.

composite score[a]--A score that combines several scores by a specified formula.

concurrent criterion-related validity[a]--Evidence of criterion-related validity in which predictor and criterion information are obtained at approximately the same time.

construct[a]--A psychological characteristic (e.g., numerical ability, spatial ability, introversion, anxiety) considered to vary or differ across individuals. A construct (sometimes called a latent variable) is not directly observable; rather it is a theoretical concept derived from research and other experience that has been constructed to explain observable behavior patterns. When test scores are interpreted by using a construct, the scores are placed in a conceptual framework.

cost accounting approach--The approach used to develop a dollar criterion that considers the value of products and services and the organization's costs to provide products and services.

cost effectiveness[c]--A state or condition in which the benefits associated with a particular outcome clearly exceed the cost of obtaining the outcome.

decision[c]--A moment of choice in an ongoing process of evaluating alternatives with a view to selecting one or some combination of them to attain the desired end.

decision tree[c]--A framework for developing the anatomy of a decision making situation that uses the concepts of probability, utility, and expected value.

decision theoretic approach--The set of alternatives, costs and possible outcomes leading to a choice.

differential validity--The level of prediction using LSEs of differences among criterion scores when referring to H_d; this measure is related to the variation of a validity vector with jobs and to an assignment variable being more valid for its own job family than any other job family.

discounting--A procedure for equating the costs and benefits that accrue over time to reflect the opportunity costs and returns foregone.

efficiency--A solution that minimizes costs as measured by physical resources and time utilized.

expected value[c]--A concept that permits a decision maker to place a monetary or other value on the positive and negative consequences likely to result from the selection of a particular alternative.

external employee movement--The analysis of employee separations and acquisitions in an organization.

goal[c]--A subset of an objective expressed in terms of one or more specific dimensions.

gross national product--The sum of all expenditures on goods and services by households, by firms on new capital, and by government.

hierarchical classification efficiency--All classification efficiency not explainable as allocation efficiency; it capitalizes on disparate variances of the mean predicted benefit scores for the corresponding jobs.

hierarchical layering--A phenomenon in which LSEs are more valid or of more value for some jobs than for others.

human capital--The skills of the workforce that determine what workers can contribute to the production process.

human resource accounting--The economic consequences of employees' behavior.

inter-rater reliability[a]--Consistency of judgments made about people or objects among raters or sets of raters.

interest inventory[a]--A set of questions or statements that is used to infer the interests, preferences, likes, and dislikes of a respondent.

inventory[a]--A questionnaire or checklist, usually in the form of a self-report, that elicits information about an individual. Inventories are not tests in the strict sense; they are most often concerned with personality characteristics, interests, attitudes, preferences, personal problems, motivation, and so forth.

item analysis[a]--The process of assessing certain characteristics of test items, usually the difficulty value, the discriminating power, and sometimes the correlation with an external criterion.

job analysis[a]--Any of several methods of identifying the tasks performed on a job or the knowledge, skills, and abilities required to perform that job.

job relatedness[b]--The inference that scores on a selection instrument are relevant to performance or other behavior on the job; job relatedness may be demonstrated by appropriate criterion-related validity coefficients or by gathering evidence of the relevance of the content of the selection instrument, or of the construct measured.

joint probability[c]--The probability that two or more events will occur.

labor--The worker effort available to the production process.

law of diminishing returns--As the quantity of an input is increased and the quantity of other inputs stays the same, a point is reached where the additional output produced per unit of added input declines.

linear combination[b]--The sum of scores, whether weighted differentially or not, on different assessments to form a single composite score.

linear model[c]--A model of choice in which the evaluation of each alternative is based on the sum of its weighted values on all its dimensions, and the alternative with the greatest sum is the obvious choice.

longitudinal study[a]--Research that involves the measurement of a single sample at several different points in time.

marginal cost--The cost of producing an additional unit.

maximizing behavior[c]--An approach to decision making oriented toward obtaining an outcome of the highest quantity or value.

mean predicted performance (MPP)--The measurement of benefits can be approximated by computing MPP across jobs; if MPP is weighted by the value of each job, it becomes a more useful measure of benefits. It provides a means of comparing the effectiveness of alternative tests or test batteries in the context of a specified set of jobs and performance scores.

meta-analysis[b]--A procedure to cumulate findings from a number of validity studies to estimate the validity of the procedure for the kinds of jobs or groups of jobs and settings included in the studies.

meta-analysis--A technique for determining the degree to which the variance in validity coefficients across situations for job-test combinations is due to statistical artifacts.

model[c]--A physical or abstract representation of some part of the real world that is used to describe, explain, or predict behavior.

Monte Carlo analysis--A stochastic technique that can provide numerical solutions for mathematical functions lacking analytic solutions; the analysis typically uses random numbers as input to an evaluation process employing variance reduction procedures.

multidimensional screening (MDS)--A selection/classification process using an algorithm that insures no nonselected person has a higher predicted performance on any job than the person assigned to that job; the algorithm also ensures that no other assignment can further raise the mean predicted performance.

multivariate[b]--Characterizing a measure or study that incorporates several variables.

norms[a]--Statistics or tabular data that summarize the test performance of specified groups, such as test takers of various ages or grades. Norms are often assumed to represent some larger population, such as test takers throughout the country.

norm-referenced test[a]--An instrument for which interpretation is based on the comparison of a test taker's performance to the performance of other people in a specified group.

objective[b]--Pertaining to scores obtained in a way that minimizes bias or error due to different observers or scores.

operational efficiency--The improvement in MPP obtained from the usually imperfect operational selection assignment process as contrasted to potential efficiency, the improvement obtainable if the maximally efficient prediction composites of a given battery were to be used in optimal selection/assignment algorithms.

opportunity cost[c]--The cost of the next best alternative that is sacrificed to select what appears to be the best alternative.

payoff[c]--The intersection of an alternative and a state of nature in a payoff table; it measures the value (utility) to the decision maker likely to result from the selection of that alternative given the probabilistic occurrence of the state of nature.

payoff table[c]--A convenient framework in which to present the elements of a decision making situation employing the concepts of probability, utility, and expected value.

percentile[a]--The score on a test below which a given percentage of scores fall.

performance[b]--The effectiveness and value of work behavior and its outcomes.

personality inventory[a]--An inventory that measures one or more characteristics that are regarded generally as psychological attributes or interpersonal skills.

placement--A procedure in which individuals are matched to levels within jobs as contrasted to the classification process of matching personnel to jobs.

potential allocation efficiency--The maximum allocation effectiveness achievable from the differential validity of a given test battery and set of jobs expressed as a mean predicted performance standard score.

potential classification efficiency--The maximum classification effectiveness achievable from a given test battery and set of jobs expressed as a mean predicted performance standard score; it incorporates both potential allocation efficiency and hierarchical layering effects.

potential selection efficiency--Rank-ordering applicants on some benefit continuum and rejecting all those below some point on that continuum.

potential utilization efficiency--The sum of potential selection efficiency and potential classification efficiency.

predictive criterion-related validity[a]--Evidence of criterion-related validity in which criterion scores are observed at a later date, for example, for job or school performance.

predictor[a]--A measurable characteristic that predicts criterion performance such as scores on a test, evidence of previous performance, and judgments of interviewers, panels, or raters.

productivity--The ratio of outputs to inputs of a resource (workers, capital equipment); a measure of the degree of the use of resources..

psychometric[a]--Pertaining to the measurement of psychological characteristics such as abilities, aptitudes, achievement, personality, traits, skill, and knowledge.

regression equation[b]--An algebraic equation used to predict criterion performance from predictor scores.

relevance[b]--The extent to which a criterion measure reflects important job performance dimensions or behaviors.

reliability[a]--The degree to which test scores are consistent, dependable, or repeatable, that is, the degree to which they are free of errors of measurement.

reliability coefficient[a]--The square of the correlation of an observed score with its "true" component; often measured as the coefficient of correlation between two administrations of a test. The conditions of administration may involve variation of test forms, raters or scorers, or passage of time. These and other changes in conditions give rise to qualifying adjectives being used to describe the particular coefficient, e.g., parallel form reliability, rater reliability, test retest reliability, etc.

residual score[a]--The difference between the observed and the true or predicted score.

restriction of range[a]--A situation in which, because of sampling restrictions, the variability of data in the sample is less than the variability in the population of interest.

risk[c]--A common state or condition in decision making characterized by the possession of incomplete information regarding a probabilistic outcome.

sample[b]--The individuals who are actually tested from among those in the population to which the procedure is to be applied.

score[a]--Any specific number resulting from the assessment of an individual; a generic term applied for convenience to such diverse measures as test scores, estimates of latent variables, production counts, absence records, course grades, ratings, and so forth.

selection--A procedure for rejecting some applicants for organizational membership as contrasted to assigning all applicants to jobs (classification); or rejecting an applicant for a single job as contrasted to selection and assignment to one of a number of jobs (multidimensional selection).

selection decision[a]--A decision to accept or reject applicants for a job on the basis of information.

selection instrument[b]--Any method or device used to evaluate characteristics of persons as a basis for accepting or rejecting applicants.

selection procedures[b]--Process of arriving at a selection decision.

sensitivity analysis--An analytic technique in which a utility parameter is varied through a range of values, holding other parameter values constant to determine the impact on the total utility estimates.

shrinkage[a]--Refers to the fact that a prediction equation based on a first sample will tend not to fit a second so well.

shrinkage correction[b]--Adjustment to the multiple correlation coefficient for the fact that the beta weights in a prediction equation cannot be expected to fit a second sample as well as the original.

simulation model[c]--A special type of abstract model that is analogous to a segment of the real world and contains a time dimension. It is used to explain and predict behavior as if it occurred in the real world.

skill[b]--Competence to perform the work required by the job.

split-half reliability coefficient[a]--An internal analysis coefficient obtained by using half the items on the test to yield one score and the other half of the items to yield a second, independent score. The correlation between the scores on these two half-tests, stepped up via the Spearman-Brown Formula, provides an estimate of the alternate-form reliability of the total test.

standard score[a]--A score that describes the location of a person's score within a set of scores in terms of its distance from the mean in standard deviation units.

standardized prediction[b]--A test employed for estimating a criterion of job performance, the test having been developed and normative information produced according to professionally prescribed methods as described in standard reference works.

standards[c]--Criteria against which the results of an implemented decision can be measured.

state of nature[c]--A state or condition likely to prevail when a choice is made.

sunk costs--Costs that once incurred cannot be changed by future action.

test[b]--A measure based on a sample of behavior.

test fairness--The most commonly accepted model of test fairness is the regression model; a fair test predicts the job performance of a minority and the majority in the same way.

test-retest coefficient[a]--A reliability coefficient obtained by administering the same test a second time to the same group after a time interval and correlating the two sets of scores.

trade-off value[c]--A value that exists when a given amount of one kind of performance may in some measure be substituted for another kind of performance.

traditional selection approach--The view of tests as measuring instruments intended to assign accurate values to attributes of an individual stressing precision of measurement and estimation rather than selection outcomes.

unidimensionality[a]--A characteristic of a test that measures only one latent variable.

utility[c]--Technically, want-satisfying power; it is often defined as the preference of the decision maker for a given outcome.

utility analysis--The determination of institutional gain or loss (outcomes) anticipated from various courses of action usually measured in terms of dollars.

validity[a]--The degree to which a certain inference from a test is appropriate or meaningful.

validity coefficient[a]--A coefficient of correlation that shows the strength of the relation between predictor and criterion.

validity generalization[a]--Applying validity evidence obtained in one or more situations to other similar situations on the basis of simultaneous estimation, meta-analysis, or synthetic validation arguments.

values[c]--The nominative standards by which human beings and organizations are influenced in their choices.

variability[b]--The spread or scatter of scores.

variable[a]--A quantity that may take on any one of a specified set of values.

variance[a]--A measure of variability; the average squared deviation from the mean; the square of the standard deviation; and, in the experimental design literature, the sum of the squared deviation from its mean doubled by the degrees of freedom.

Z-score[a]--A type of standard score scale in which the mean equals zero and the standard deviation equals one unit for the group used in defining the scale.

NOTES:

a Adapted from American Psychological Association, American Educational Research Association, and National Council on Measurement in Education (1985). *Standards for Education and Psychological Testing.*

b Adapted from Society for Industrial and Organization Psychology (1987). *Principles for the Validation and Use of Personnel Selection Procedures.*

c Adapted from Heyne (1988). *Microeconomics.*

REFERENCES

Anastasi, A. (1988). *Psychological Testing* (6th ed.). New York: Macmillan.

Armor, D.J., Fernandez, R., Bers, R., and Schwarzbach, D. (1982, September). *Recruit aptitudes and Army job performance: Setting enlistment standards for infantrymen.* R-2874-MRAL. Santa Monica, CA: The Rand Corporation.

Baldwin, R.H. and Dault, T.V. (1984). The cost of high quality recruits. *Armed Forces and Society*, **II**, 96-114.

Belgrave, F., and Nogami, G. (1986, August). *Variables related to attrition in the Army.* Paper presented at the meeting of the Association of Black Psychologists Convention, Oakland, CA.

Berryman, S., Bell., R., and Lisowski, W. (1983). *The military enlistment process: What happens and can it be improved?* R-2986-MRAL. Santa Monica, CA: The Rand Corporation.

Blair, J.D., and Phillips, R.L. (1983). Job satisfaction among youth in military and civilian work settings. *Armed Forces and Society*, **9**, 555-658.

Boudreau, J.W. (1983). Economic considerations in estimating the utility of human resource productivity improvement programs. *Personnel Psychology*, **36**, 551-576.

Brogden, H.E. (1949). When testing pays off. *Personnel Psychology*, **2**, 171-183.

Brogden, H.E. (1951). Increased efficiency of selection resulting from replacement of a single predictor with several differential predictors. *Educational and Psychological Measurement*, **11**, 173-196.

Brogden, H.E. (1959). Efficiency of classification as a function of number of jobs, percent rejected, and the validity and intercorrelation of job performance estimates. *Educational and Psychological Measurement*, **19**, 181-190.

Buddin, R. (1981). *The role of service experience in post-training attrition in the Army and Air Force.* R-2682-MRA. Santa Monica, CA: The Rand Corporation.

Buddin, R. (1984). *Analysis of Early Military Attrition Behavior.* R-3069-MIL. Santa Monica, CA: The Rand Corporation.

Cardinet, J. (1959). The use of profiles for differential classification. *Education and Psychological Measurement*, **XIX**, 191-205.

Cascio, W.F. (1987). *Costing human resources: The financial impact of behavior in organizations* (2nd ed.). Boston: PWS-Kent.

Christal, R.E. (1968, Summer). Selecting a harem, and other applications of the policy-capturing model. *Journal of Experimental Education*, **36**, 35-41.

Commission on the All Volunteer Force (1970). *Studies for the Commission on the All Volunteer Armed Forces. Volume 2.* Washington, DC: U.S. Government Printing Office.

Congressional Budget Office (1986). *Quality soldiers: Costs of manning the active Army.* Washington, DC: Author.

Cronbach, L.J., and Gleser, G.C. (1965). *Psychological tests and personnel decisions.* Second Edition. Urbana: University of Illinois Press. Originally published 1957.

Daula, T. and Smith, D. (1984). Recruiting goals, enlistment supply, and enlistments in the U.S. Army. In C. Gilroy (Ed.), *Army Manpower Economics* (Chapter 3). Boulder, CO: Westview Press.

Dertouzos, J.N. (1985). *Recruiter incentives and enlistment supply.* R-2065-MIL. Santa Monica, CA: The Rand Corporation.

Eaton, N.K. (1987, March). *The Army's Project A, improving selection, classification, and utilization of Army enlisted personnel.* Briefing for HQDA Personnel Proponent General Officer Steering Committee Conference. Alexandria, VA: U.S. Army Research Institute.

Eaton, N.K., Wing, H., and Mitchell, K.J. (1985). Alternate method of estimating the dollar value of performance. *Personnel Psychology*, **38**, 27-40.

Eaton, N.K., and Nogami, G. (1981). Army male and female attrition. In W. Sinaiko (Ed.), *Military personnel attrition and retention research in progress* (pp. 45-56). Washington, DC: Office of Naval Research.

Eiger, A., Jacobs, J., Chung, D., and Selsor, J. (1988). The U.S. Army's occupational specialty manpower decision support system. *Interfaces*, **18**, 57-73.

Fernandez, R., and Garfinkle, J. (1985). *Setting enlistment standards and matching recruits to jobs using job performance criteria.* R-3067-MIL. Santa Monica, CA: The Rand Corporation.

Flyer, E.S. (December, 1959). *Factors relating to discharge for unsuitability among 1956 airman accessions to the Air Force.* Technical Note WADC-TN-59-201. Dayton, OH: Wright Air Development Center.

Goodstadt, B.E. and Glickman, A.S. (November, 1975). *The current status of enlisted attrition in the U.S. Navy and in the U.S. Marine Corps and the search for remedies.* AIR-54500-11/75-FR. Washington, DC: American Institute for Research.

Guion, R.L. (1965). *Personnel testing.* New York: McGraw-Hill.

Harris, R.N. (1967, March). *A model sampling experiment to evaluate the methods of test selection.* Research Memorandum 67-2, Statistical Research and Analysis Division. Washington, DC: U.S. Army Behavior and Systems Research Laboratory.

Hillier, F.S. and Lieberman, G.J. (1967). *Introduction to Operations Research.* San Francisco, CA: Holden-Day, Inc.

Horne, D. (1985). Modeling Army enlistment supply for the All Volunteer Force. *Monthly Labor Review*, **108**, 35-39.

Horst, P. (1954). A technique for the development of a differential prediction battery. *Psychological Monographs*, **68** (9, Whole No. 380).

Hunter, J.E. (1983). *Validity generalization for 12,000 jobs: An application of job classification and validity generalization analysis to the General Aptitude Test Battery (GATB).* USES Test Research Report No. 45. Washington, DC: U.S. Employment Service, U.S. Department of Labor.

Hunter, J.E. (1984). *The validity of the Armed Forces Vocational Aptitude Battery (ASVAB) High School Composite.* Report for Research Applications, Inc., in partial fulfillment of DoD Contract No. F41689-83-C-0025.

Hunter, J.E. (1985). *Differential validity across jobs in the military.* Report for Research Applications, Inc., in partial fulfillment of DoD Contract No. F41689-83-C-0025.

Hunter, J.E., Crosson, J.J., and Friedman, D.H. (1985). *The validity of the Armed Services Vocational Aptitude Battery for civilian and military job performance.* Rockville, MD: Research Applications, Inc.

Hunter, J.E., and Schmidt, F.L. (1982). Fitting people to jobs: The impact of personnel selection on national productivity. In E.A. Fleishman and M.D. Dunnette (Eds.), *Human performance and productivity: Vol. 1: Human capability assessment.* Hillsdale, NJ: Erlbaum.

Johnson, C.D. and Zeidner, J. (1989, August). Classification utility: Measuring and improving benefits in matching personnel to jobs. IDA Paper P-2240. Alexandria, VA: Institute for Defense Analyses.

Kroeker, L., and B. Rafacz (1983). *Classification and assignment within PRIDE (CLASP): A recruit assignment model.* Technical Report 84 9. San Diego, CA: Navy Personnel Research and Development Center.

Lawshe, C.H. (1952). What can industrial psychology do for small business? *Personnel Psychology*, **5**, 31-34.

Maier, M.H. and Fuchs, E.F. (1972, September). Development and evaluation of a new ACB and aptitude area system. Technical Research Note 239. Alexandria, VA: U.S. Army Research Institute.

Maier, M.H. and Grafton, F.C. (1981, May). *Aptitude composites for ASVAB 8, 9, and 10.* Research Report 1308. Alexandria, VA: U.S. Army Research Institute.

Maier, M., and Truss, A. (1983). *Original scaling of ASVAB forms 5, 6, and 7: What went wrong.* CRC 457. Alexandria, VA: Center for Naval Analyses.

Manganaris, A.G., and Schmitz, E.J. (1984). *Projecting attrition by military occupational specialty.* Technical Report 637. Alexandria, VA: U.S. Army Research Institute.

Manganaris, A.G., and Schmitz, E.J. (1985). *Impact of delayed entry program participation on first term attrition.* Technical Report 665. Alexandria, VA: U.S. Army Research Institute.

Mason, R.O., and Flamholtz, E.G. (1978). Human Resource Management. In S. Moder and S.E. Elmaghraby (Eds.) *Handbook of Operations Research,* Vol. 2. New York, NY: Van Nostrand Reinhold Co. (pp. 92-110).

McCormich, E.J., Jeanneret, P.R., and Mecham, R.C. (1972). A study of job characteristics and job dimensions as based on the Position Analysis Questionnaire. *Journal of Applied Psychology Monograph* **56**, 347-368.

McHenry, J.J. (1987, April). *Project A validity results: The relationship between predictor and criterion domains.* Paper presented at the Society for Industrial and Organizational Psychology Annual Conference, Atlanta, GA.

McLaughlin, D.H., Rossmeissl, P.G., Wise, L.L., Brant, D.A., and Wang, M.M. (1984, October). *Validation of current and alternative ASVAB area composites, based on training and SQT information of FY81 and FY82 enlisted accessions.* Technical Report 651. Alexandria, VA: U.S. Army Research Institute.

Mossholder, K.W. and Arvey, R.D. (1984). Synthetic Validity: a conceptual and comparative review. *Journal of Applied Psychology*, 69, 322-333.

Nelson, A. and Schmitz, E. (1985) Multiobjective personnel allocation. In *Revolution and Evolution: New Directions in OR/MS* (E. Partridge, Ed.) Washington, DC: Washington Operations Research/Management Science Council.

Nelson, A., Schmitz, E.J., and Promisel, D. (1984). *The impact of personnel quality on Stinger weapon system performance.* Technical Report 640. Alexandria, VA: U.S. Army Research Institute.

Nicholson, W. (1978). *Microeconomic Theory.* Hinsdale, IL: The Dryden Press.

Niehaus, R.J. (1979). *Computer-assisted human resource planning.* New York: Wiley.

Nord, R.D. and White, L.A. (1988, May). *Performance utility and optimal job assignment.* Draft Technical Report. Alexandria, VA: U.S. Army Research Institute.

Polich, J., Dertouzos, J., and Press, S. (1986). *The enlistment bonus experiment.* R-3353-FMP. Santa Monica, CA: The Rand Corporation.

Primoff, E.S. (1975). *How to prepare and conduct job element examinations.* Washington, DC: U.S. Civil Service Commission, Personnel Research and Development Center.

Sadacca, R., Campbell, J.P., White, L.A., and DiFazio, A.S. (1988, July). *Weighting criterion components to develop composite measures of job performance.* HumRRO IR-PRD-88-14. Alexandria, VA: Human Resources Research Organization.

Sadacca, R., White, L., Campbell, J.P., DiFazio, A.S., and Schultz, S.R. (1989). *Assessing the utility of MOS performance levels in Army enlisted occupations*. Technical Report 839. Alexandria, VA: U.S. Army Research Institute.

Sands, W.A. (1978). Enlisted personnel selection for the U.S. Navy. *Personnel Psychology*, **31**, 63-70.

Schmidt, F.L., and Hunter, J.E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, **62**, 529-540.

Schmidt. F.L., and Hunter, J.E. (1981). Employment testing: Old theories and new research findings. *American Psychologist*, **36**, 1128-1137.

Schmidt, F.L., Hunter, J.E., and Dunn, W.L. (1987, November). *Potential utility increases from adding new test to the Armed Services Vocational Aptitude Battery (ASVAB)*. Contract No. Delivery Order 0053. San Diego, CA: U.S. Navy Personnel Research and Development Center.

Schmidt, F.L., Hunter, J.E., and Larson, M. (1988, August). *General cognitive ability vs. general and specific aptitudes in the prediction of training performance: Some preliminary findings*. Contract No. Delivery Order 0053. San Diego, CA: U.S. Navy Personnel Research and Development Center.

Schmidt, F.L., Hunter, J.E., McKenzie, R.C., and Muldrow, T.W. (1979). The impact of valid selection procedures on workforce productivity. *Journal of Applied Psychology*, **64**, 609-626.

Schmidt, F.L., Hunter, J.E., Pearlman, K., and Shane, G.S. (1979). Further tests of the Schmidt-Hunter validity generalization model. *Personnel Pscyhology*, **32**, 257-281.

Schmitz, E. and McWhite, P. (1986). *Evaluating the benefits and costs of the enlisted personnel allocation system*. Technical Report 721. Alexandria, VA: U.S. Army Research Institute.

Schmitz, E. and Nelson, A. (1984a). *The allocation of personnel to military occupational specialties*. Technical Report 635. Alexandria, VA: U.S. Army Research Institute.

Schmitz, E. and Nelson, A. (1984b). *Modeling the enlistment process for personnel planning*. Research Report 1416. Alexandria, VA: U.S. Army Research Institute.

Schmitz, E. and Nord. R. (1987). *Evaluating improvements to MOS assignment policy*. MPPRG WP87-X. Alexandria, VA: U.S. Army Research Institute.

Scribner, B., Smith, D., Baldwin, R., and Phillips, R. (1986). Are smart tankers better? AFQT and military productivity. *Armed Forces and Society*, **12**, 193-206.

Tamaki, M. (1984). The secretary problem with optimal assignment. *Operations Research*, **32**, 847-858.

Trattner, M.H. (1982). Synthetic validity and its application to the Uniform Guidelines validation requirements. *Personnel Psychology*, **35**, 383-397.

Varian, H.R. (1978). *Microeconomic analysis*. New York, NY: W.W. Norton and Company, Inc.

Ward, J.H. (1977, August). *Creating mathematical models of judgment processes: From policy capturing to policy specifying*. AFHRL-TR-77-47. AD-AO48983. Brooks AFB, TX: Air Force Human Resources Laboratory.

Ward, J., Haney, D., Hendrix, W., and Pina, M. (1978). *Assignment procedures in the Air Force procurement management information system*. Technical Report 78-30. Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Ward, J.E., Pina, M., Fast, J.C., and Roberts, D.K. (1979, October). *Policy specifying with application to personnel classification and assignment*. Paper presented at Military Testing Association Meeting, San Diego, CA.

Wise, L.L., Campbell, J.P., and Arabian, J.U. (1988). The Army synthetic validation project. In B.F. Green, H. Wing, and A.K. Wigdor (Eds.), *Linking military enlistment standards to job performance* (pp. 76-85). Committee on the Performance of Military Personel Commission on Behavioral and Social Sciences and Education. National Research Council. Washington, DC: National Academy Press.

Wise, L.L., Campbell, J.P., McHenry, J.J., and Hanser, L.R. (1986, August). *A latent structure model of job performance factors*. Paper presented at the American Psychological Association Annual Meeting, Washington, DC.

Wise, L.L., Campbell, J.P., and Peterson, N.G. (1987, April). *Identifying optimal predictor composites and testing for generalizibility across jobs and performance constructs*. Paper presented at the Second Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Zeidner, J. (1987, April). *The validity of selection and classification procedures for predicting job performance*. IDA Paper P-1977. Alexandria, VA: Institute for Defense Analyses.

Zeidner, J. and Johnson, C.D. (1989, July). Selection utility for military and civilian jobs. IDA Paper P-2239. Alexandria, VA: Institute for Defense Analyses.